REMARKS

This amendment and reply is in response to the Office Action dated November 3, 2006, and the Advisory Action dated May 17, 2007. According to the Advisory Action, Applicants' reply mailed May 3, 2007, does not place the application in condition for allowance because "[t]he arguments presented have already been addressed in the Final Rejection." Applicant disagrees that the arguments presented had been addressed in the Final Rejection, and respectfully request that the Examiner consider the amendments and remarks herein. The amendment of the claims and remarks herein are identical to the amendment and remarks in the reply mailed May 3, 2007. In addition, Applicant has noted the remarks that had not been addressed in the Final Rejection mailed November 3, 2006 ("the Final Rejection").

Claims 1, 8, 18, and 26 have been amended. Claims 1, 8, 18, and 26 have been amended to delete reference to SEQ ID NOs:1 and 2. Claim 1 has been amended to replace "85%" with "95%" and to refer to specific regions of SEQ ID NO:3. Support for this amendment is found throughout the specification as filed, e.g., in original claim 8, and page 2, lines 34-38. Claim 26 has been amended to refer to insert the phrase "consisting of the fully complementary sequence of" SEQ ID NO:3. This amendment is supported by claim 26 as originally filed. Claims 6 and 23 have been canceled.

A replacement Sequence Listing is enclosed herewith. The replacement Sequence Listing includes the polypeptide sequences depicted in the drawings filed with the application.

No new matter has been added by the present amendment. Upon entry of this amendment, claims 1, 6, 8, 15-18, and 26, will be pending and under examination.

Rejections under 35 U.S.C. § 101

Utility

Claims 1 (c), 6, 8 (g-i), 15-17, 18 (c), 23, and 26 (c) were rejected as lacking patentable utility. In the Office Action, it was alleged that

> [t]he polypeptide encoded by a nucleic acid comprising the nucleotide sequence of SEQ ID NO:3 lacks a specific utility...The instant specification is silent as to

> the function and proteolytic breakdown of polypeptides encoded by SEQ ID NO:3, having at least 85% identity to SEQ ID NO:3, or having at least 30-100 nucleotides from SEQ ID NO:3. Indeed, no polypeptides or polypeptides [sic] sequences are taught in the specification (page 3).

This is respectfully traversed. The claims, as amended, are drawn to various polypeptides, e.g., encoded by a nucleic acid molecule at least 95% identical to a polypeptide encoded by nucleotides 585-2156 of SEQ ID NO:3, nucleotides 2307-5741 of SEQ ID NO:3, or nucleotides 5620-7533 of SEQ ID NO:3; to polypeptides encoded by sequences at least 30 nucleotides in length from SEQ ID NO:3; and to polypeptides encoded by a sequence at least 100 nucleotides in length that hybridizes under stringent conditions to SEQ ID NO:3. Applicant disagrees with the statement that "no polypeptides" or polypeptide sequences are taught in the specification. The specification discloses nucleotide sequences and describes regions of the sequences that encode retroviral polypeptides (e.g., gag, pol, and env polypeptides). The specification also discloses polypeptide sequences encoded by the nucleotide sequences. See, e.g., Figures 2 and 3.

Applicant maintains that the asserted utilities are specific. A utility is specific if it is specific to the subject matter claimed. Applicant has asserted utilities for the claimed polypeptides, for example, in methods of detecting porcine retroviruses and screening animals for retrovirus infection. See the specification, e.g., at page 24, lines 23-28, page 30, lines 25-27, and page 39, lines 13-20. A utility is general if it is applicable to the broad class of the invention. The asserted utilities do not apply to *any* polypeptide. These cannot be characterized as general, non-specific utilities. One cannot carry out methods for detecting swine retroviruses with any, non-specific polypeptide. Applicant has indicated that the claimed polypeptides are useful for particular reasons.

Furthermore, the references cited by the Examiner regarding characterization of proteins of other retroviruses do not establish lack of utility for the claimed polypeptides. In fact, these references indicate *well-established* utilities for the claimed polypeptides. A well-established utility is a specific, substantial, and credible utility which is well known, immediately apparent, or implied by the specification's disclosure of the properties of a material, alone or taken with the knowledge of one of skill in the art. MPEP 2107.02.II.B. The claimed polypeptides are

products of a mammalian retrovirus. It is acknowledged in the Office Action that "SEQ ID NO:3 encodes at least 3 different proteins, gag, pol, and env" (page 3). The polypeptide components of retroviruses have a number of conserved, well-known functions. As noted in Freed (*Virology*, 251:1-15, 1998), "[t]he Gag proteins of HIV-1, *like those of other retroviruses*, are necessary and sufficient for the assembly of virus-like particles" (abstract, first sentence, emphasis added). Products of the pol region encodes reverse transcriptase and integrase enzymes (See, e.g., Coffin, *Fund. Virol.*, Ch. 27, page 654, 1991, copy enclosed herewith as Exhibit A). Env gene products are expressed on the surface of virions and mediate interactions with cell-surface receptors (Coffin, *Fund. Virol.*, Ch. 27, page 654, 1991). Certain functions of retroviral polypeptides are well known to those of skill in the art. The foregoing remarks were not addressed in the Final Rejection.

The Office Action cited lack of details regarding proteolysis of the claimed polypeptides as evidence of lack of utility. The relevance of this to the well-established and asserted utilities is not understood. Freed, cited in the Office Action, says that "[e]xpression of retroviral Gag precursor proteins is both necessary and sufficient for the assembly and release of noninfectious, virus-like particles; Gag processing by [viral protease] is *not required* for particle production" (*J. Virol.*, 76(10):4679-4687, 2002, at page 4679, left col., second paragraph, emphasis added). Thus, proteolysis is not required for some functions of retroviral proteins. The foregoing remarks were not addressed in the Final Rejection.

Regarding Applicant's discussion of asserted utilities in the prior Amendment in Reply, filed September 1, 2006, the Office Action stated that

> [t]his argument is not persuasive because this is a "reach through" utility, that is, one must make all of the polypeptides encoded by SEQ ID NO:3, make all of the antibodies specific to these polypeptides, then determine for themselves if the detection of these polypeptides has any bearing on xenograft transfer. It is not for another to arrive at Applicant's invention. As noted, many polypeptides are encoded by SEQ ID NO:3. Applicants have not provided a function for any of these polypeptides…One cannot know until they determine for themselves if detection of any one of the polypeptides will be an indication that a donor animal will pass the nucleic acid retroviral vector to the xenograft recipient and cause deleterious effects. Thus, the polypeptides encoded by SEQ ID NO:3 lack a specific utility (page 5).

This is respectfully traversed. The asserted utilities are not "reach through" utilities. Applicant has asserted that the claimed polypeptides are useful, e.g., in methods of detecting retroviruses for screening animals and organs for transplantation. The Office Action cites no factual basis for concluding that one needs to "make all of the polypeptides encoded by SEQ ID NO:3... then determine for themselves if the detection of these polypeptides has any bearing on xenograft transfer" in order to practice the asserted utilities. The asserted utilities are not ones that require further research to identify or reasonably confirm. One of skill in the art possesses the technical capabilities to produce and use the claimed polypeptides. Applicant has explained that swine are a potential source of organs for xenotransplantation (specification, page 1, lines 16-17). Detection and analysis of swine viruses is important for recognizing infection in swine and managing infection in xenotransplantation.

Applicant notes that an assertion of utility is credible unless (A) the logic underlying the assertion is seriously flawed, or (B) the facts upon which the assertion is based are inconsistent with the logic underlying the assertion. MPEP 2107.2.III.B. The Office Action suggests that the utilities lack credibility where it states that "[o]ne cannot know until they determine for themselves if detection of any one of the polypeptides" will be an indication that a donor animal will transmit a retrovirus and cause deleterious effects. If this rejection is maintained, Applicant requests clarification as to what is meant by this statement. Is it doubted whether positive detection of a retrovirus protein correlates with virus transmission? In the prior Amendment in Reply, Applicant cited evidence that swine retroviruses infect human cells. Is it doubted whether transmission of a retrovirus causes disease? Given the knowledge of the involvement of retroviruses in disease, not to mention the desire to avoid unnecessary exposure to pathogens, particularly in immunocompromised transplant recipients, the logical basis for screening animals and tissues to avoid retrovirus transmission is neither "seriously flawed" nor based on inconsistent facts. The foregoing remarks were not addressed in the Final Rejection. In view of the foregoing, Applicant requests withdrawal of this rejection.

Rejections under 35 U.S.C. § 112, first paragraph

Claims 1(c), 6, 8 (g-i), 15-17, 18(c), 23, and 26(c) were rejected as lacking written description. According to the Office Action,

> [t]he specification and claims do not set forth any structure of or function for the claimed polypeptides encoded by SEQ ID NO:3, having at least 85% identity to SEQ ID NO:3, or having at least 30 or 100 nucleotides from SEQ ID NO:3. Also, this polypeptide is not in hand. (Office Action, page 6).

This rejection is respectfully traversed. The claims are directed to polypeptides encoded by SEQ ID NO:3, polypeptides encoded by nucleic acid molecules that have a high degree of identity to SEQ ID NO:3, polypeptides encoded by nucleic acid molecules that hybridize under high stringency conditions to a molecule consisting of the fully complementary sequence of SEQ ID NO:3. The following remarks, and the amendments to claims 8, 18, and 26, were not addressed in the Final Rejection.

The specification discloses SEQ ID NO:3 and polypeptides encoded by SEQ ID NO:3. See Figure 3. The statement that the specification does not set forth "any" structure for the claimed polypeptides is incorrect. The specification provides complete structure for polypeptides that fall within the claims. For example, the specification discloses the complete amino acid sequence of polypeptides encoded by nucleotides 585-2156 of SEQ ID NO:3, nucleotides 2307-5741 of SEQ ID NO:3, and nucleotides 5620-7533 of SEQ ID NO:3. See Figure 3. The genera of polypeptides encompassed by the claims are related by a high degree of identity to polypeptides encoded by SEQ ID NO:3. Possession of a claimed invention can be shown in a number of ways. Recitation of structure is one of ways to demonstrate possession. The correlation of structure with function can be used to support written description when minimal structure is disclosed. In the present case, far more than minimal structure is disclosed. The claims are limited to polypeptides encoded by sequences related by a high degree of identity to SEQ ID NO:3. The structural disclosure satisfies the written description requirement.

Applicant maintains that one of skill in the art is able to identify polypeptide sequences encoded by a given nucleotide sequence. The Examiner rejected this argument because "there is no correlation of structure with function" (Office Action, page 6). One need not know the function of a polypeptide *a priori* to be able to identify an open reading frame in a nucleotide sequence. It is noted that claim 1, as amended, and claim 8, refer to the specific region(s) of SEQ ID NO:3 that encode(s) the polypeptide.

Applicant requests withdrawal of the rejection of claims 1(c), 6, 8 (g-i), 15-17, 18(c), and 26(c) as allegedly lacking written description.

Rejections under 35 U.S.C. § 112, second paragraph

Claims 1(c), 6, 8 (g-i), 15-17, 18(c), 23, and 26(c) were rejected as indefinite for referring to "a polypeptide." The Office Action stated that "retroviruses encode many polypeptides; therefore it is not clear which polypeptide is being claimed" (page 6).

This rejection is traversed. Claims 1, 8, 18, and 26 are genus claims, i.e., they encompass multiple species.[1] Genus claims are acceptable ways to claim one's invention. MPEP 2173.05(h).

Claim 1 was rejected for reciting "85% identical." The Office Action stated that "'identical' is an absolute term, meaning that one thing is identical to another or it is not. Thus, one skilled in the art cannot know what a fraction of identical means" (page 7).

This is traversed. The meaning of a claim term is the meaning that the term would have to a person of ordinary skill in the art in question at the time of the invention. The meaning of a term in the claims may be evidenced by the specification. The term "identical," in the context of nucleotide and amino acid relatedness, has an art-recognized meaning, which is also set forth in the specification at page 25, lines 23-33. It is indeed possible for polymeric macromolecules to be identical at some positions and mismatched at others when the sequences of the macromolecules are aligned. The interpretation of "identical" as a purely absolute term is inconsistent with its usage in the art and the specification. Despite the Examiner's interpretation of "identical" as an absolute term, percent identity was used by those of skill in the art at the time the application was filed to describe the degree of relatedness of sequences. Two abstracts that use this terminology are enclosed. These published in 1986 and 1990. Both refer to percent identity to describe relatedness of polypeptide and nucleotide sequences.

[1] Claims 6 and 23 have been canceled.

Claim 18 was rejected for reciting "70% homology." The Office Action stated that "'homology' is a qualitative term and not a quantitative term. Thus, one skilled in the art cannot know what 70% homology means" (page 7).

This is traversed. "Homologous," in the context of nucleotide and amino acid relatedness, is expressly defined in the specification at page 25, lines 23-33. Understood in light of the knowledge of one of skill in the art and the definition in the specification, the term "homology" is clear.

Claim 18 was also rejected for using the term "corresponding." According to the Office Action, "it is not clear what a corresponding human, mouse, or primate retrovirus sequence is, or the last five 3' bases may be" (page 7).

This is traversed. Applicant request that the Examiner consider the point made previously in the Amendment in Reply filed September 1, 2006. Retroviruses have a common genomic structural organization. A region is corresponding if it is at the same relative position. One of skill in the art can compare a sequence from one region of a swine retrovirus to a sequence in a region at the same relative position in a human, mouse, or primate retrovirus. For example, a sequence at the 5' end of a gag gene of a swine retrovirus can be compared to a sequence at the 5' end of a gag gene of a human retrovirus.

Claim 18 uses the phrase, "wherein the last five 3' nucleotides are unique to the selected sequence." This phrase is also clear to one of skill in the art. It simply refers to the five nucleotides at the 3' end of the sequence of SEQ ID NO:3 encoding the polypeptide. To satisfy the limitation, the last five base pairs of the sequence from SEQ ID NO:3 must be unique, i.e., the base pairs are not identical to the base pairs in the corresponding human, mouse, or primate retroviral sequence.

Claim 18 was rejected for lacking "(c)" preceding a limitation in the claim. This rejection is moot in view of the amendment to the claim to remove lettered subheadings.

The foregoing remarks, and amendments to the claims, were not addressed in the Final Rejection.

Non-elected inventions have been canceled from the claims under examination.

In view of the foregoing, Applicant requests withdrawal of the rejections of the claims under 35 U.S.C. § 112, second paragraph.

Objection to the Specification

The specification was objected to for failing to include the polypeptide sequences depicted in the drawings in the Sequence Listing. A substitute Sequence Listing is being filed herewith which includes the polypeptide sequences.

A Request for Continued Examination, Petition for Extension of Time, and required fees are being filed herewith. Please apply any other charges or credits to deposit account 06-1050, referencing attorney docket no. 14846-011004.

Respectfully submitted,

Date: Sept. 4, 2007

Laurie Butler Lawrence
Reg. No. 46,593

Fish & Richardson P.C.
225 Franklin Street
Boston, MA 02110
Telephone: (617) 542-5070
Facsimile: (617) 542-8906

Fundamental Virology, Second Edition, edited by B. N. Fields, D. M. Knipe et al. Raven Press, Ltd., New York © 1991

CHAPTER 27

Retroviridae and Their Replication

John M. Coffin

GROUP DEFINITION

No group of viruses has received as much attention from scientists in recent years as retroviruses. The intense scrutiny given these agents not only reflects their importance as human and animal pathogens but also reflects their remarkable value as experimental objects, which, in turn, is a consequence of their intimate association with the host. The unique replication cycle of these viruses leads (directly or indirectly) to a variety of special biological features, which include:

1. The existence—within the constraints of a common virion structure, genetic organization, and replication cycle—of a large number of virus strains with very different lifestyles and pathogenic effects.
2. A wide variety of interactions between virus and host, ranging from the completely benign infections displayed by endogenous viruses through moderate exogenously acquired infections to the generally fatal consequences of viruses such as HIV and the rapidly oncogenic viruses.
3. The ability to acquire and alter the structure and function of host-derived sequences and, as a consequence, to present them to the researcher as oncogenes, the study of which has provided our most fundamental insights into molecular mechanisms of carcinogenesis.
4. The ability to insert themselves into the germline of a host and behave as a transposable element, with genetic consequences that have been an important force in vertebrate evolution.
5. The ability to cause certain types of genetic damage, such as activation or inactivation of specific genes near the site of integration of the provirus.
6. The ability to rapidly alter their genomes by mutation and recombination in response to altered environmental conditions.
7. The ability to serve as vectors for foreign genes inserted in the laboratory and to carry and express these genes in a predictable way into a wide variety of cells and organisms.

In this chapter, I will describe the general features

J. M. Coffin: Department of Molecular Biology and Microbiology, Tufts University, Boston, Massachusetts 02111.

Fundamental Virology, Second Edition, edited by B. N. Fields, D. M. Knipe et al.
Raven Press, Ltd., New York © 1991

CHAPTER 27

Retroviridae and Their Replication

John M. Coffin

GROUP DEFINITION

No group of viruses has received as much attention from scientists in recent years as retroviruses. The intense scrutiny given these agents not only reflects their importance as human and animal pathogens but also reflects their remarkable value as experimental objects, which, in turn, is a consequence of their intimate association with the host. The unique replication cycle of these viruses leads (directly or indirectly) to a variety of special biological features, which include:

1. The existence—within the constraints of a common virion structure, genetic organization, and replication cycle—of a large number of virus strains with very different lifestyles and pathogenic effects.
2. A wide variety of interactions between virus and host, ranging from the completely benign infections displayed by endogenous viruses through moderate exogenously acquired infections to the generally fatal consequences of viruses such as HIV and the rapidly oncogenic viruses.
3. The ability to acquire and alter the structure and function of host-derived sequences and, as a consequence, to present them to the researcher as oncogenes, the study of which has provided our most fundamental insights into molecular mechanisms of carcinogenesis.
4. The ability to insert themselves into the germline of a host and behave as a transposable element, with genetic consequences that have been an important force in vertebrate evolution.
5. The ability to cause certain types of genetic damage, such as activation or inactivation of specific genes near the site of integration of the provirus.
6. The ability to rapidly alter their genomes by mutation and recombination in response to altered environmental conditions.
7. The ability to serve as vectors for foreign genes inserted in the laboratory and to carry and express these genes in a predictable way into a wide variety of cells and organisms.

In this chapter, I will describe the general features

J. M. Coffin: Department of Molecular Biology and Microbiology, Tufts University, Boston, Massachusetts 02111.

of retrovirus biology, as illuminated by study of all virus groups used more or less interchangeably. Special features of the human retroviruses, the human T-cell lymphoma viruses (HTLVs), and the human immunodeficiency virus (HIV) are covered in detail in their own chapters and will be discussed only in passing here. Similarly, retroviral oncogenesis is thoroughly discussed in Chapter 13 and will only be touched upon here.

Even with these limitations, it is impossible to cover every well-studied aspect of retrovirology in any significant depth. The two-volume *RNA Tumor Viruses* (461,462), which is about 2,500 pages long, is still current in many respects and should be consulted by the reader interested in more information, as should a number of recent reviews on specialized aspects of the subject (25,28,30,53,64,135,156,245,294,308,383,385, 430,440–442,444,472). Because of the very large size of the retroviral literature, I will make no attempt to be comprehensive in citations. In particular, references to papers prior to the appearance of the second volume of *RNA Tumor Viruses* in 1985 will be minimal, and the reader is again encouraged to consult "the bible."

The retroviridae comprise a large family of viruses, primarily of vertebrates, although there have been a few reported sightings in other animals. Both in the wild and in the laboratory, they are associated with many diseases, including rapid and long-latency malignancies, wasting diseases, neurological disorders, and immunodeficiencies, as well as lifelong viremia in the absence of any obvious ill effects. Despite the variety of interactions with the host, all retrovirus isolates are quite similar in virion structure, genome organization, and mode of replication.

The virion is enveloped and is about 100 nm in diameter. Its surface is decorated by a single protein structure, probably a trimer of two protein subunits, products of the *env* gene. The internal nucleocapsid, or core, is an ill-defined, roughly spherical to conical structure made up of the three or four products of the *gag* gene. Also included in the core are several proteins that have important catalytic roles during replication. These include a protease and two products of the *pol* gene: the reverse transcriptase whose several enzymatic activities conspire to convert the genetic information from single-strand RNA to double-strand DNA, and integrase, necessary for covalently joining virus to cell DNA to form the provirus.

The genome consists of two (usually identical) molecules of single-strand RNA, ranging from about 7 to 10 kb in length, modified in ways reminiscent of cell mRNAs, including capping at the 5′ end and polyadenylation at the 3′ end. The order of the genes encoding structural proteins is invariably *gag–pol–env*. A number of other genes involved in regulation of virus expression are present in some virus groups.

The replication cycle can be thought of as proceeding in two phases. The first phase includes (a) entry of the virion core into the cytoplasm, (b) synthesis of double-strand DNA using the single-strand genome as template, (c) transfer of the structure to the nucleus, and (d) integration of the DNA into the host genome. These steps are mediated by proteins found within the virion and proceed in the absence of viral gene expression. The second phase includes (a) synthesis and (b) processing of viral genomes, mRNAs, and proteins using host-cell systems (such as RNA polymerase), sometimes aided by the presence of specific viral gene products. Virion assembly proceeds by encapsidation of the genome by gag and gag–pol fusion proteins, processing of these to the finished products, association of the nucleocapsids with the cell membrane, and release of the virion by budding.

CLASSIFICATION

Retroviruses have been traditionally divided into three subfamilies, based primarily on pathogenicity rather than on genome relationships (423,424) (see Chapter 2). Viruses are further described according to the following: (a) virion structure (types A to D and others; see later); (b) utilization of particular cell receptors; (c) lifestyle, whether endogenous (i.e., passed from parent to offspring as a provirus integrated into the germline) or exogenous; (d) presence or absence of an oncogene; and (e) other pathogenic properties. When nucleotide sequence relationship and genome structure are used as criteria, seven groups of viruses comprising all well-analyzed isolates can be recognized. These groups and some well-known strains are listed in Table 1. (The names assigned to them are based on routine usage and should be considered provisional.)

Viruses originally isolated as tumor-inducing agents, as well as related viruses, are traditionally placed into the subfamily Oncovirinae. Since this subfamily includes five groups that are not closely related to one another, these will be described and referred to as separate entities.

The Avian Leukosis-Sarcoma Virus (ALSV) Group

This group includes both exogenous and closely related endogenous viruses of birds. Viruses of this group have C-type virions and genomes which encode only virion structural genes (*gag*, *pol*, *env*), although many isolates of exogenous virus are further modified by the presence of oncogenes, such as *src* in Rous sarcoma virus (RSV). They are further divided ac-

TABLE 1. *Retrovirus groups*

Subfamily	Group	Example isolates	Comments
Oncovirinae	Avian leukosis-sarcoma	Rous sarcoma virus (RSV)	Exogenous; oncogene containing (*src*)
		Avian myeloblastosis virus (AMV)	Exogenous; oncogene-containing (*myb*)
		Avian erythroblastosis virus (AEV)	Exogenous; oncogene-containing (*erb*-A and *erb*-B)
		Rous-associated virus (RAV)-1 to 50	Exogenous; cause B-lymphoma, osteopetrosis, and other diseases
		RAV-0	Endogenous; benign
	Mammalian C-type	Moloney murine leukemia virus (Mo-MLV)	Exogenous; causes T-cell lymphoma
		Harvey murine sarcoma virus (Ha-MSV)	Exogenous; oncogene-containing (*H-ras*)
		Abelson murine leukemia virus (A-MuLV)	Exogenous; oncogene-containing (*abl*)
		AKR-MuLV	Endogenous; benign
		Feline leukemia virus (FeLV)	Exogenous; causes T-cell lymphoma, immunodeficiency, and many other diseases
		Simian sarcoma virus	Exogenous; oncogene-containing (*sis*) (SSV)
		Numerous endogenous and exogenous viruses, mostly in mammals	
		Reticuloendotheliosis virus (REV); spleen necrosis virus (SNV)	Exogenous viruses of birds
	B-type viruses	Mouse mammary tumor virus (MMTV)	Endogenous and exogenous; mostly milk-borne; causes mostly mammary carcinoma, some T-lymphoma
	D-type viruses	Mason–Pfizer monkey virus (MPMV)	Exogenous; unknown pathogenicity
		"SAIDS" viruses	Immunodeficiencies in monkeys
	HTLV–BLV group	Human T-cell leukemia (or lymphotropic) virus (HTLV)	Causes T-cell lymphoma; associated with neurological disorders
		Bovine leukemia virus (BLV)	Causes B-cell lymphoma
Lentivirinae	Lentiviruses	Human immunodeficiency virus (HIV-1 and -2)	Cause of AIDS
		Simian immunodeficiency virus (SIV)	Causes AIDS-like disease in certain monkeys
		Feline immunodeficiency virus (FIV)	
		Visna/maedi virus	Causes neurological and lung disease in sheep
		Equine infectious anemia virus (EIAV)	
		Caprine arthritis-encephalitis virus (CAEV)	
Spumavirinae	"Foamy" viruses	Many human and primate isolates [e.g., simian foamy virus (SFV)]	Exogenous, apparently benign

cording to host range (i.e., receptor utilization) into seven subgroups, denoted A through G. The first four subgroups are characteristic of exogenous viruses from chickens; subgroup E belong to endogenous viruses of chickens; and subgroups F and G are from endogenous viruses of pheasants.

The Mammalian C-Type Virus Group

These viruses include a large number of endogenous and exogenous viruses and are represented by isolates from many groups of mammals, including rodents, carnivores, and primates, as well as some exogenous viruses of birds (i.e., the reticuloendotheliosis virus group). The human genome is sprinkled with closely related defective endogenous proviruses, but no replicating human viruses of this group have been isolated. Like the avian viruses, which they superficially resemble, nondefective mammalian C-type viruses contain only *gag*, *pol*, and *env* genes, and many oncogene-containing isolates are known. Viruses of mice and cats can be further classified by host range. The murine viruses are (unfortunately) designated by the

species distribution of their receptors: *ecotropic* viruses replicate only in mouse cells; *xenotropic* viruses use receptors found on cells of most species except mice; and *polytropic* and *amphotropic* viruses use different receptors found in both mouse and nonmurine species. Exogenous murine viruses are ecotropic or amphotropic; endogenous viruses are found with eco, poly, or xenotropic host ranges. Feline leukemia viruses are arbitrarily classified into subgroups A through C, with endogenous viruses belonging to subgroup C.

The B-Type Virus Group

This group includes as infectious agents only the mouse mammary tumor virus (MMTV), isolated as both endogenous and exogenous (but vertically transmitted) viruses. Related but defective sequences can be found in the genomes of many species, including rats, monkeys, and humans. These viruses contain the *gag*, *pol*, and *env* genes as well as an additional coding region, ORF, of undefined function. No oncogene-containing variants have been described.

The D-Type Virus Group

This group includes exogenous isolates from primates, such as the Mason–Pfizer virus, isolated from a mammary carcinoma of a rhesus monkey but of uncertain pathogenicity. More recent isolates include a virus associated with an immunodeficiency syndrome (SAIDS) in some captive monkey colonies. The genomes of these viruses seem to contain only the three virion protein genes (394). No oncogene-containing isolates have been described.

The HTLV–BLV Group

This group includes exogenous viruses associated with B-cell lymphoma in cattle and with T-cell lymphoma, as well as with some neurological diseases in humans. No endogenous relatives or oncogene-containing viruses of this group are known. In addition to genes encoding virion proteins, these viruses encode at least two nonvirion proteins important for gene expression.

Lentivirinae

This subfamily includes exogenous viruses responsible for a variety of neurological and immunological diseases, but not directly implicated in any malignancies. The prototype members of this family were the "slow" viruses visna, equine infectious anemia, and caprine arthritis–encephalitis. More recently, it has become clear that HIV and the related simian and feline immunodeficiency viruses also belong to this group. Genomes of these viruses are characterized by a complex combination of genes in addition to *gag*, *pol*, and *env*.

Spumavirinae

The foamy viruses are by far the least well characterized of the retroviruses. They have been isolated as agents that cause vacuolation ("foaming") of cells in culture from a number of mammalian species, including monkeys, cattle, cats, and humans. Persistent infection with these viruses is not associated with any known disease. From the partial nucleotide sequence of the genome of one isolate (269), it is clear that they do not belong to any of the other groups.

Evolutionary Relationships

Analysis of the nucleotide sequence of a large number of retrovirus isolates has yielded relationship trees similar to the one shown in Fig. 1 (88,194,261,473). Although it is impossible to affix a uniform time scale to such a tree, and attempts to do so (387) should not be taken seriously, there is no reason to believe it does not accurately reflect the relationships among the groups. These relationships provide a basis for the grouping presented here: Between groups the relationship is barely visible, even in the most highly conserved regions of the genome; within a group the relationship of genome structure and sequence is quite obvious.

VIRION STRUCTURE

Types of Particles

Retroviruses are united by a common virion structure, shown in a highly schematic cartoon form in Fig. 2. Some of the earliest studies using the electron microscope to probe the virion and its synthesis revealed some differences in detail (which could be used to distinguish among the various groups) and led to the classification into four morphological groups (types A to D), a classification still used to some extent today (423,424), although it is not current with the more recently described viruses.

A-type particles were first viewed as the immature intracellular forms of MMTV, and the term is still reserved for strictly intracellular structures. Virtually indistinguishable structures are also found in cell lines derived from a number of rodent species (mouse my-

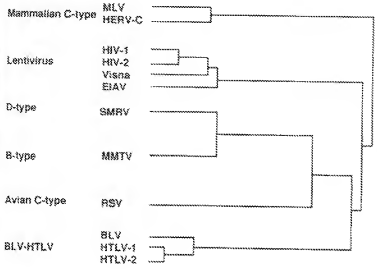

FIG. 1. Relationships of retrovirus groups. The relationships are based on amino acid sequence similarities in the RT protein of the groups shown (89,195,271). Note that the scale is approximate and not necessarily linear. HERV-C is an ancient endogenous provirus of humans (347); SMRV is squirrel monkey retrovirus; other abbreviations are defined in Table 1.

elomas are a famous example). These are the product of a fairly high copy number endogenous provirus-like element known as IAP (for intracisternal A particle) (60,411). As the name implies, this form is usually seen within intracellular membranes, although apparently similar cytoplasmic forms can also be seen. A-type particles are hollow, roughly spherical structures that are about 60–90 nm in diameter with a double-walled appearance.

B-type particles are the enveloped, extracellular form of MMTV. Freshly budded forms resemble enveloped A-type particles, but they rapidly mature to a form with a tightly condensed, acentric nucleocapsid. MMTV virions are also characterized by prominent surface spikes, comprising the env protein.

C-type particles characterize most of the avian and mammalian oncoviruses studied to date, despite the lack of close sequence relationship among the various groups. Typically, complete intracellular forms are not observed; rather, the first visible forms are crescent-shaped patches at the site of budding from the cell membrane. Budding and assembly seem to occur simultaneously; nearly complete immature forms are within nearly completed envelopes. As with B-type particles, immature, freshly budded forms display hollow nucleocapsids; with maturation comes condensation of the core into a central, electron-dense structure. This morphological change is a consequence of cleavage of the gag precursor protein (see later). C-type particles generally have barely visible surface projections, significantly less prominent than those of B-type viruses.

D-type particles are defined only in the primate viruses such as MPMV and the SAIDS isolates. They resemble B-type viruses, having a complete intracellular nucleocapsid and an eccentric location of the core in mature particles. They differ in that they have less prominent surface projections.

Other virion types, not as yet assigned a letter, are seen in the other virus groups. BLV and the HTLVs resemble C-type viruses in their mode of budding and in having a central nucleocapsid, but they differ in the appearance of the envelope proteins. Lentiviruses, including HIV, also bud like C-type viruses, but the ma-

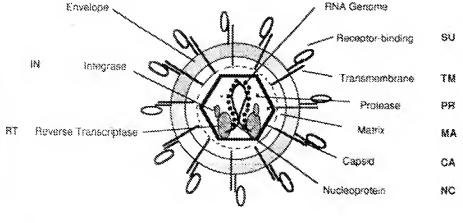

FIG. 2. The retrovirus virion. This highly schematic figure shows the relative locations of the various structures and proteins. It is not intended to be an accurate depiction of their organization.

ture virion has a distinctive bar (or truncated cone)-shaped nucleocapsid. Spumaviruses display (a) intracellular A-particle-like cores and (b) virions with prominent surface projections.

Genome Organization

The retrovirus genome is unique among viruses in several aspects, including its physical organization, its mode of synthesis, and its functions in replication. It is (a) the only diploid genome, (b) the only one synthesized and processed by the cell's mRNA-processing machinery, (c) the only one to be associated with a specific RNA whose function is to prime replication, and (d) the only positive-strand RNA genome that does not serve as mRNA early after infection.

Physical Structure

Figure 3 is a diagram illustrating the physical features common to all retrovirus genomes. These invariably include (a) modifications to the RNA imposed by the cellular machinery responsible for its synthesis and

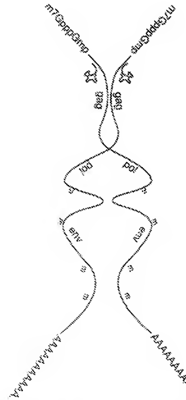

FIG. 3. Features of the *retrovirus* genome. The dimeric genome includes the following, from 5' to 3': the m^7 Gppp capping group, the primer tRNA, the coding regions, the m_6A residues (m), and the 3' poly(A) sequence.

processing and (b) noncovalent arrangements imposed on the genome during virion assembly.

A Capping Group

In all retrovirus genomes analyzed to date, this modification, applied to the 5' end during synthesis by the cellular transcription machinery, takes the form of m7G5'ppp5'G_mp. This structure is presumably important for the translation of those molecules which serve as messages. Its role, if any, in replication is unknown.

Internal Methylation

Retroviral genomes, like cell mRNAs, are also modified posttranscriptionally by methylation on the 6 position of occasional A residues. Careful analysis of the sites of methylation on the RSV genome reveals that the methylation is somewhat sequence-specific (200), with the favored sequence of the form . . . PuGm^6ACU . . . Not all such sites are modified, however, and it has been suggested that the 6-methylation might be important for the regulation of splicing (408).

Polyadenylation

All retrovirus genomes contain a string of about 200 A residues at their 3' end, a modification typical of most eukaryotic mRNAs. Again the poly(A) sequence is added as a posttranscriptional modification to newly made transcripts by the cell's mRNA processing machinery. Most genomes contain a canonical signal of the form AAUAAA within about 20 bases of the site of poly(A) addition. A notable exception to this is the genome of viruses in the HTLV-BLV group, in which the nearest such signal is more than 250 bases upstream of the poly(A) site (437). As with normal cell mRNA, the AAUAAA sequence, perhaps in combination with less well characterized signals (31,92), specifies cleavage of a longer precursor RNA and, in a coupled reaction, subsequent poly(A) addition. It has not been directly determined whether poly(A) addition is a necessary feature of retrovirus genomes or simply reflects their mode of synthesis, although indirect evidence implies the former (92). The site of polyadenylation seems relatively unimportant. Mutations that partly or completely inactivate the ALSV polyadenylation signal have only a small effect on genome packaging and replication, even though most or all of the genomes are elongated into cell sequences at their 3' ends and are polyadenylated (162,417).

Dimer Structure

If isolated from virions under nondenaturing conditions, retrovirus genomes have physical properties

(sedimentation and electrophoretic mobility) consistent with a molecule about twice the size of the subunit size determined for denatured molecules. It was once believed that the genome subunits were different in sequence (as would be the case with a number of other RNA viruses, such as influenza), but careful determination of the sequence complexity of the genome revealed that the two molecules must have identical sequence (i.e., the genome must be diploid).

The structure of the complex is very poorly understood. Complicated models predicting secondary structure have been presented; some invoke participation of the primer tRNA in the structure (61). It is not yet clear how seriously these should be taken. The two genomes must be associated at numerous points along their length; however, titration of the complexes with denaturing agent preparatory to examination in the electron microscope reveals the most stable joining point to be near the 5′ end of each genome. As yet, the physical reality of these awaits a critical test.

The role of diploidy in the retrovirus lifestyle remains to be determined. While it has been proposed that viral DNA synthesis makes use of the two RNA molecules to synthesize one of DNA (322), there is no obvious necessity for this in our current model of reverse transcription (see later). One clear consequence of having two genomes per virion is the extraordinarily high rate of recombination seen in retroviruses, and it may well be that this recombination frequency permits the effective repair of physical damage to the genome (see later).

Association with Other RNA Molecules

Denaturation of native genomes releases a number of small RNA molecules—largely tRNAs and small ribosomal RNAs. The most important of these is a single molecule per genome of tRNA which is associated via base pairing of its 3′-terminal 18 nucleotides to a complementary sequence located 100–500 bases from the 5′ end of the genome. This tRNA molecule serves a key role in replication: It is the primer on which reverse transcriptase initiates DNA synthesis. The specific tRNA that is appropriated by the virus for this role varies from group to group. At present, four tRNAs are known which serve this function (Table 2). In general, the specific tRNA primer used is highly conserved within a retrovirus group, but among different MLVs two different tRNAs (pro and glu) have been found to be used (67,411).

Other Virion Nucleic Acids

In addition to the genome and associated RNAs, other nucleic acid molecules can be found in retrovirus virions. These include small RNAs (a variety of tRNAs, 5S ribosomal RNAs, and others), traces of mRNA (such as globin), and even some DNA (461). Although they are not merely a random representation of cellular nucleic acids, there is presently no good reason to ascribe functional significance to these inclusions in the virion. It should be borne in mind, however, that such molecules might, as a rare event, be copied into DNA that could be subsequently integrated into the cell genome and thus give rise to processed pseudogenes or other genomic rearrangements (89,242).

Cis-Acting Regions

Roughly speaking, retrovirus genomes are arranged so that all (or almost all) noncoding sequences that contain important recognition signals for DNA or RNA synthesis and processing are located in terminal regions, with internal regions given over virtually entirely to protein coding functions. The general sequence organization of a typical retrovirus genome is displayed in Fig. 4. Reading from the 5′ end, the important noncoding features common to all retroviruses include the following:

R. All retrovirus genomes are terminally redundant,

TABLE 2. *Terminal regions of retrovirus genomes*[a]

Virus group	Approximate size (in bases)					Primer tRNA
	U3	R	U5	Leader	3′ Untrans	
ASLV	150–250	18–21	80	250	150	Trp
Mam C-type	500	60	75	475	40	Pro/Glu
B-type	1,200	15	120	130	[b]	Lys-3
D-type	240 +	15 +	95 +	145	100	Lys-1,2
HTLV–BLV	250–350	120–240	100–200	50–100	[c]	Pro
Lenti	450	100	60–80	150	[b]	Lys-1,2
Spuma	800	200	150	90	[b]	Lys-1,2

[a] Data were taken from refs. 116, 267, 283, 337, 393, 394, and 437.
[b] The 3′ open-reading frame overlaps U3

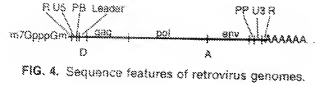

FIG. 4. Sequence features of retrovirus genomes.

containing the identical sequence at the 5′ end, adjacent to the capping group and at the 3′ end immediately preceding the poly(A) tract. This sequence plays an important role during reverse transcription in permitting the transfer of nascent DNA from one end of the genome to the other. The R sequence varies considerably in length, from a low of 12 bases in MMTV to a current record high of 235 in BLV.

U5. This region, containing unique information near the 5′ end of the genome, is defined by its flanking sequences—R and the primer-binding site. U5 is the first region copied into DNA during reverse transcription, and it becomes the 3′ end of the LTR. Mutational analysis implies multiple roles of this region. Some U5 sequences are essential for initiation of reverse transcription (58,293). The 3′ end contains one of the *att* sites necessary for integration, and there is evidence for a role of this region in packaging of viral RNA (293).

PB. The 18 nucleotides that form the primer-binding (PB) site are invariably perfectly complementary to the 3′-terminal nucleotides of the tRNA primer. The length of the PB region is defined by the position of an uncopyable base in the tRNA—an m^1 A residue (133).

Leader. Like some other RNA viruses, retroviruses have unusually long sequences preceding the first translated region. The untranslated sequence between the primer-binding site and the beginning of *gag* has at least two functional roles. Usually, it contains the donor site for the generation of all spliced subgenomic mRNAs encoded by the virus. In many viruses, there is only one such message known—that encoding the env proteins. Even in the highly complex sets of spliced mRNAs displayed by lentiviruses, all subgenomic species are spliced from this donor, with variation in the downstream site used and the presence of additional splices distinguishing the individual species.

The other important function encoded in the leader region is to specify incorporation of genome RNA into virions (61,62,245). The so-called packaging signal has been roughly localized by analysis of spontaneous or introduced mutations (208,244,265,303). In at least one case—Moloney MuLV—it seems to extend into gag (19), and insertion of a sequence encompassing this region is sufficient to confer the ability to be packaged into virions upon unrelated RNAs (1). In most retroviruses, the location of the packaging signal downstream of the splice donor site ensures its absence from the viral mRNAs and consequently prevents incorporation of these into virions. In viruses of the ALSV group, by contrast, important signals for packaging seem to be largely upstream of the splice donor site near the beginning of *gag*; how mRNAs are excluded from virions is not well understood.

Internal signals. Sequences acting in *cis* are mostly found at the ends of the genome. In the majority of retroviruses, the sequence from the beginning of *gag* to the end of *env* is translated in its entirety; there is no internal nontranslated region. Indeed, there is some overlapping of reading frames. Internal *cis*-acting signals are primarily the splice acceptor sites used to form the various mRNAs. A few exceptions are notable: In lentiviruses and HTLV, complex internal regulatory sequences apparently interact with systems specified by both virus and cell to regulate relative concentrations of mRNA for structural and regulatory proteins (156,441). In ALSV, the DNA form of a sequence within *gag* confers enhancer activity when added to standard assay plasmids (7,48,202). In general, however, sequences required in *cis* for virus replication are not found interspersed with coding regions. This division, which allows replication of retrovirus genomes in which coding sequences are deleted or replaced by virtually any sequence (such as an oncogene), is of central importance to development of retrovirus vectors.

The 3′ untranslated region. In most oncoviruses, there is a short and rather variable untranslated sequence separating the end of *env* and the beginning of U3. This region is sometimes (incorrectly) referred to as the "c" region. In the ALSV group, mutations in this region diminish incorporation of genomes into virions (397). Some transcriptional enhancer activity has been ascribed to a portion of this sequence (226); the function of such an activity is uncertain at best.

The polypurine tract. All retrovirus genomes contain a characteristic sequence—a run of at least nine A and G residues—immediately preceding the beginning of U3. This sequence contains the initiation site for synthesis of the positive strand of viral DNA; apparently it provides a cleavage site for a reverse-transcriptase-associated endonucleolytic activity (114,133).

U3. The U3 region is defined as the region between the initiation site of positive-strand DNA synthesis and the beginning of R. It forms the 5′ portion of the LTR, and, in the DNA form, contains a number of *cis*-acting signals necessary for virus replication. The sequence at the immediate 3′ end of U3 is a signal recognized by the integration machinery, and it is an approximate inverted copy of the matching signal in U5. Since U3 ends at the 5′ end of the provirus, it also contains signals that are recognized by the cellular transcription machinery and that are responsible for much or all of the transcriptional control observed. Near its 3′ end

are other transcriptionally important signals, the canonical consensus sequences that mark most eukaryotic promoters. In some groups (those with short R regions; see Table 2) the consensus polyadenylation signal is also found in the 3′ portion of U3.

R. At the 3′ end of the genome, between U3 and the poly(A) sequence, is the other copy of the R region, which sometimes contains the poly(A) addition signal.

Coding Regions

Figure 5 shows the distribution of protein-coding sequences in the genomes of the various retrovirus groups. Considerable variation is apparent within a common framework. All viruses have the genes encoding virion proteins arranged in a common order, but each group has special features that distinguish it from the rest. The common features are discussed first. Special features that characterize the HTLV and lentivirus groups will be discussed only briefly. Complete details can be found in the appropriate chapters.

The gag Region

The 5′-most gene of all retrovirus genomes is named in honor of the first recognition of the proteins encoded by it as group-specific antigens. The *gag* gene is translated from the full-length RNA to yield a precursor polyprotein that is subsequently cleaved to yield three to five capsid proteins. The three invariant proteins

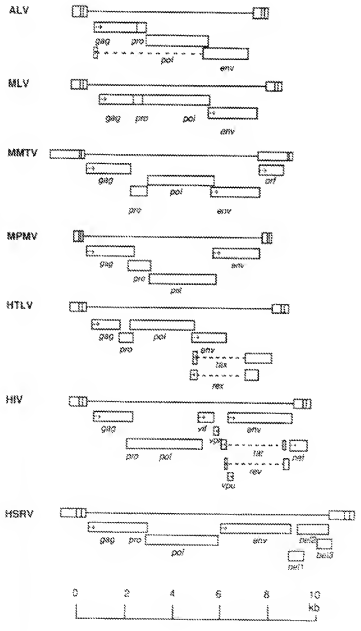

FIG. 5. Coding regions of retrovirus genomes. Open-reading frames in the genes of representative members of each retrovirus group are shown relative to the proviral structure. Small arrows show sites of translation initiation. Note the different orientation of *pro* and *pol* relative to *gag*, and also note the overlap of *gag* and *env* in most viruses. Sequences were taken from refs. 267, 283, 394, and 437. The human spumavirus (HSRV) reading frames are predicted from the sequence but have not been verified.

are (in their order of translation): the matrix (MA) protein, the capsid (CA) protein, and the nucleic-acid-binding protein (NC). In the avian and mammalian C-type viruses, an additional cleavage product, found in virions but of unknown function, occupies the region between the MA and CA peptides. The AUG initiation codon for *gag* is about 300–500 bases from the 5′ end and is rarely the first potential initiation codon. Rather, *gag* is often preceded by short open-reading frames, which have been suggested to play a role in alignment of ribosomes (148). In some mammalian C-type viruses, occasional use of an additional initiation codon in frame with, and upstream of, *gag* leads to synthesis of a cell-surface variant of the *gag* protein (86).

The pro Region

The *pro* gene encodes the protease (PR) responsible for the cleavage of the *gag* and *pol* polyproteins. Strictly speaking, neither *pro* nor *pol* should be considered as distinct genes, since they are translated only as carboxy-terminal extensions of a fraction of the *gag* proteins, and all precursor proteins which contain them also contain the entire *gag* precursor. The *pro* region is variously arranged with respect to the *gag* and *pol* regions that flank it, and the various devices employed by different groups of viruses to express it are described in a later section.

The pol Region

The *pol* gene encodes the two proteins containing the two activities needed by the virus early in infection: the reverse transcriptase (RT) and the IN protein needed for integration of viral into cell DNA. In ALSV, there is an additional short, apparently nonfunctional cleavage product encoded at the 3′ end of *pol* (5,207). As with *pro*, translation of *pol* into protein is made possible only by an occasional slip of the translation machinery which causes the preceding termination codon to be bypassed and enables translation to continue in the pol reading frame. In most retrovirus groups, with the exception of lentiviruses, the 3′ end of pol overlaps the beginning of env in a different reading frame, so there is no intervening untranslated sequence.

The env Region

The *env* gene encodes the two envelope glycoproteins that are themselves cleaved from a larger precursor. The larger of the two, the surface (SU) protein, is responsible for recognition of cell-surface receptors, and the smaller transmembrane (TM) protein anchors the complex to the virion envelope. Unlike *gag*, *pro*, and *pol*, *env* is translated from a spliced subgenomic RNA, and a characteristic splice acceptor sequence can always be identified in the appropriate location upstream of the initiation codon.

Other Genes

While many viruses (such as the avian and mammalian C-type virus groups) require only the *gag*, *pro*, *pol*, and *env* genes and their products to fill all virus-coded roles necessary for replication, other virus groups encode additional proteins that play other roles in replication. Our understanding of the functions and products of these genes ranges from none to primitive, but in general they seem to actively modify the rate and pattern of gene expression of the virus that encodes them. A number of these proteins fall under the classification of "*trans*-activators," analogous to those encoded by many DNA viruses. Other functions have also been identified. The standard nomenclature recently published for the human retroviruses and their relatives (128) will be followed throughout. Outdated synonyms are in parentheses.

B-Type Viruses

MMTV was the first retrovirus to reveal an additional gene, located 3′ of env and extending well into U3. A glycosylated protein product of molecular weight 37,000, along with the spliced mRNA that encodes it, have been sighted in cells expressing the MMTV genome (340). It has been proposed to have *trans*-activating activity (38), but variant viruses lacking much of this region can also be found *in vivo* under certain circumstances (10,94,176,230,277). The lack of attributable function accounts for the provisional name *orf*, standing for "open-reading frame."

HTLV and BLV

These viruses have a long region between *env* and U3, which encodes at least two (and possibly more) proteins in overlapping open-reading frames. This region was originally called "X" or "lor" (for long open-reading frame). Both genes are translated from a doubly spliced mRNA, which contains the initiation codon and a few additional codons from *env*. These genes include the following:

tax (p40*x*, *x-lor*, *tat*-1,2). As suggested by the name, the protein encoded by *tax* is the HTLV *trans*-activator derived from the X region. The presence of the 40,000-

dalton protein encoded by it is essential for any LTR-driven expression of these viruses.

rex (p27x). The name stands for regulator from the X region. The *rex* gene encodes an approximately 27,000-dalton protein, which is essential for the expression of full-length and *env* mRNAs. In its absence, only mRNAs encoding tax and rex are found in infected cells.

Lentiviruses

HIV and its relatives hold the unchallenged retroviral record for the number of genes identified. At least nine are known to encode products, although clearly defined functions are still lacking for some. This group of viruses is unique in having an untranslated region between *pol* and *env*, which encodes at least part of several proteins.

tat (*tat*-III). This gene is named for its apparent *trans*-activating function, which greatly stimulates expression from the HIV LTR, although probably by a mechanism including posttranscriptional effects, quite different from that of *tax*. It is also quite different in size and contains only about 90 amino acids. Despite its small size, the translated portion is divided over two regions of the genome, and it is translated from a multiply spliced mRNA.

rev (*art*, *trs*). This small protein is a regulator of virus expression. It is expressed in a fashion like *tat* from an overlapping reading frame. Although very different in size and structure, it has a function similar to the HTLV rex protein in that the pattern of spliced mRNA is strongly affected by it; and, indeed, under certain conditions, one protein can substitute for the other in virus replication.

nef (3′-*orf*). This protein is encoded in an open-reading frame located 3′ of env and extending into U3. It is named for its apparent function as a "negative factor" for virus replication. In its presence, transcription driven by the HIV LTR is somewhat reduced (3). Some virus isolates have mutations in this region which cause them to not encode functional protein (294). This has the apparent effect of giving these viruses a growth advantage in cell culture over their wild-type parents.

vif (*sor*). The product of a short open-reading frame overlapping pol encodes a protein necessary for virion infectivity (114).

vpr, **vpx**. The virus protein encoded by reading frames r and x have been detected in cells infected with HIV-1 and HIV-2, respectively, but little is known about their function.

Spumaviruses

As with HTLV and BLV, the nucleotide sequence of a human spumavirus genome reveals the presence of several overlapping open-reading frames between env and U3 (116,267). The existence and function of the proteins encoded by them has not been demonstrated.

The Capsid

The structure of the nucleocapsid of retroviruses is not well understood. Our primary source of knowledge has been genetic analysis supplemented with some information gleaned from fractionation studies. When virions are carefully treated with mild detergent, two rather fragile structures are formed, depending on virus and conditions (86). The nucleoprotein contains the virus RNA, the NC protein, and a fraction of the RT activity. In addition, the core contains the CA protein. Electron microscopy of whole or fractionated virions has not been particularly revealing of the fine structure: The particles are somewhat pleomorphic and do not reveal obvious symmetries.

The majority of the structural features observed are attributable to gag proteins. Indeed, almost-normal-looking virions can be formed in the absence of pol and env proteins as well as genome RNA. The major gag proteins are most readily studied as the cleavage products found in virions. It should be borne in mind, however, that they act in two different forms at different times. During virion assembly they are unified as a precursor protein; however, it is the cleavage products that enter the cell to initiate the infection cycle. As a consequence of the mode of assembly, it can be expected that all proteins will be present in equal numbers in the virion. This expectation seems to be borne out (at least approximately), and it has been estimated (for ALSV) that there are about 2,000–4,000 copies of each cleavage product per virion (83).

Conventionally, virion proteins are now designated by a two-letter mnemonic suggesting their function (233). Traditionally, they were named according to their apparent molecular weights, by affixing a prefix (p, gp, pp, or Pr for protein, glycoprotein, phosphoprotein, or precursor, respectively) to the molecular weight in thousands, and they appear this way in much of the current literature. The correspondence between the two naming systems among the various groups is shown in Table 3. The organization of the gag, pro, and pol proteins along their respective precursors is shown in Fig. 6.

The NC Protein

The NC protein is a small basic protein found in the virion in association with the genome RNA. It is usu-

TABLE 3. *Virion proteins of retroviruses*[a]

Current name	ALSV	Mammalian C-type	MMTV	D-type	HTLV–BLV	Lentiviruses
MA	p19	p15	p10	p10	p19/15	p17
?	p10	p12	p21	p18	NP[b]	NP[b]
CA	p27	p30	p27	p27	p24	p24
NC	p12	p10	p14	p14	p12	p15
PR	p15	p14	p13	——	p14	p14
RT	p68	p80	——	——	——	——
IN	p32	p46	——	——	——	——
SU	gp85	gp70	gp52	gp70	gp60	gp46
TM	gp37	p15E	gp36	gp22	gp30	gp21

[a] The order is 5′ to 3′ from top to bottom. Data were taken from ref. 233.
[b] Not present.

ally phosphorylated on serine residues, and it has been suggested that the phosphorylation state of this protein may be an important determinant of viral assembly or disassembly (122). When the structure of this protein is compared among the different groups (and among some retrotransposons as well), a conserved feature emerges, in addition to the overall basic nature of the protein. This is a sequence of the form Cys-x_3-Cys-x_4-His-x_4-Cys, present twice in NC protein of most retrovirus groups but occurring only once in most mammalian C-type viruses (20). These "Cys-His boxes" resemble the so-called "zinc finger" domains that are well known in some DNA-binding proteins, such as the gene 32 protein of bacteriophage T4, known to form specific sites for divalent ion binding (20); moreover, binding of zinc *in vitro* to at least one NC protein has been observed (367). In virions of another retrovirus, however, NC-associated zinc has not been detected even in a very sensitive analysis (J. Leis, *personal communication*).

The CA Protein

The CA protein is hydrophobic and forms the major internal structural feature of the virion—the core shell. Beyond this, little is known about its structure or mechanism of action. Mutations in CA eliminate virus assembly altogether (177). The sequence of this protein is not highly conserved among retrovirus groups.

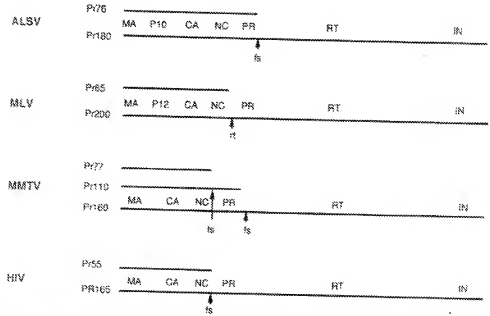

FIG. 6. Expression of the *gag*, *pro*, and *pol* genes in various retroviruses. The disposition of the coding regions for the final cleavage products is shown on each of the precursor molecules. Note the use of readthrough suppression (rt) or frameshift suppression (fs) to accomplish the synthesis of the nested products.

The MA Protein

The MA protein is the gag protein in closest association with the membrane. It is the only protein that can be chemically cross-linked to lipid in virions (328). [The idea that this protein is also a specific RNA-binding protein (85) seems to have been generally discarded (272).] Consistent with its membrane association, the amino terminus of most MA proteins is modified by the addition of a fatty acid (invariably myristic acid) group, a modification characteristic of many proteins that lie on the internal face of the cell membrane (85). Interestingly, the ALSV MA protein has only an added acetyl group at its amino terminus, and it also lacks any significant stretch of hydrophobic amino acids that might provide for membrane attachment. Although it is known to be associated with the membrane through a sequence near its amino terminus (328), the mechanism by which it associates with the membrane is not known. Other modifications that might expedite the association have not been avidly sought.

Other gag Proteins

Two retrovirus groups, the ALSV group and the mammalian C-type viruses, encode additional gag cleavage products, called p10 and p12, respectively. In both cases, this protein occupies the portion of the precursor between the MA and CA domains, and it is found in the corresponding location in virions [i.e., just outside the core shell (87)]. To date, no function has been assigned to these proteins.

Virion Enzymes

The products of the *pro* and *pol* regions constitute the enzymatic activities found in retrovirus virions. Their mode of synthesis, as occasional extension products of gag, results in their presence in much smaller numbers—less than a hundred as compared to thousands of copies per virion.

The RT Protein

The primary product of *pol* is cleaved by the virion protease to yield the amino-terminal RT peptide [which contains the activities necessary for DNA synthesis (RNA and DNA-directed DNA polymerase, ribonuclease H, and some other nucleolytic activities)] and the carboxy-terminal IN (or "integrase") protein. In the case of ALSV, the cleavage between the two occurs in only a fraction of molecules, and this results in two proteins with RT activity. The 572-amino-acid form is called α; the larger, uncleaved molecule is called β. The ALSV RT is most commonly present as an αβ heterodimer, although monomeric and homodimeric forms are also present. The activities of each are not well sorted out.

A fraction of the RT activity found in virions is in association with the viral nucleoprotein structure, and seems to be bound specifically to the primer tRNA, at least in ALSV. This binding may be important in virion assembly; ALSV mutants lacking pol are also devoid of primer tRNA (333). In MLV, specific primer-tRNA–RT binding has not been demonstrated.

The two best-studied activities of RT are the DNA polymerase and ribonuclease H. Fortunately, the retrovirus enzyme (unlike most other virion polymerases) is readily separable from the virion, and several have been successfully produced in popular cloning systems in bacteria and insect cells. Both the ALSV and MLV enzymes are commercially available. The DNA polymerase activity displayed by RT uses RNA and DNA as templates more or less interchangeably and, like all DNA polymerases known, is unable to initiate DNA synthesis *de novo* but, instead, requires a preexisting molecule to serve as primer. In the test tube, either RNA or DNA will do; in the natural case, it is always RNA: The tRNA is used for synthesis of the first (negative) strand, and the genome itself is nicked within the polypurine tract for the synthesis of the second (positive) strand. RT activity is readily detected within disrupted virions. In the absence of added template, the endogenous activity uses the genome–primer combination to synthesize negative-strand DNA. Addition of an appropriate hetero- or homopolymeric template–primer pair to disrupted virus leads to a substantially greater synthesis of the complementary DNA. The use of an RNA homopolymer template and a DNA oligonucleotide primer provides some (although far from absolute) specificity relative to adventitious cellular DNA polymerases. Poly(rC)•oligo(dG) seems to be best in this regard. This fairly sensitive assay is still used routinely in the laboratory to detect retrovirus infections in cell culture, to quantitate the amount of virus produced, and to partially characterize the virus. Many recently discovered retroviruses were first seen by virtue of their RT activity (461,472).

All RTs have an absolute requirement for divalent cation for polymerase activity *in vitro*. In all groups but one, the preferred ion is Mg^{2+} at about 10 mM. The exceptional group is the mammalian C-type viruses, whose RT prefers Mn^{2+} at about 3 mM, especially for use on homopolymer templates. This difference is occasionally useful for classification.

Early in replication the RNase H activity inherent in all RT proteins plays the essential role of removing the RNA genome as DNA synthesis proceeds (see below). It selectively degrades the RNA from an RNA–DNA hybrid molecule. On model substrates it

behaves as both exonuclease and endonuclease. On the one hand, it usually requires a base-paired RNA end; it will not cleave molecules with overhanging single-strand regions, and it cleaves progressively from either a 5′ or 3′ end. On the other hand, the cleavage products are small oligonucleotides, and it makes specific endonucleolytic cuts at strategic points in the process of DNA synthesis.

An extensive mutational analysis of the MLV RT region has revealed that the polymerase and ribonuclease H activities occupy separate, nonoverlapping domains, with the polymerase covering approximately the entire amino-terminal two-thirds of the molecule (Fig. 6) (422). This sequence alignment is consistent with that predicted from the rather distant amino acid sequence relationships to known enzymes (194).

The RT is one of the most highly conserved proteins encoded by the retroviral genome. Although the overall relationship from one virus group to the next is not close, there is enough amino acid sequence similarity between the various enzymes to detect a clear relationship, to infer the family tree shown in Fig. 1, and to discern certain key structural features. Some of these features are in common with the much more distantly related RTs of retrotransposons of flies, yeast, and plants, hepadnaviruses (Chapter 38), and cauliflower mosaic virus (88,269,440) as well as even the recently discovered enzyme in certain bacteria species (183,226,241,443). A noteworthy characteristic, for example, is the presence of a highly conserved Tyr-x-Asp-Asp sequence that may be associated with the catalytic site.

The IN Protein

The IN protein is the smaller, carboxy-terminal cleavage product of the pol protein. [In ALSV, a small, apparently nonfunctional peptide is removed from the carboxy terminus of IN during processing (5,207).] Genetic analysis provides clear evidence for the role of IN in replication. Mutations that affect its structure block virus replication (165); in some viruses, such mutations have been shown to affect integration but not DNA synthesis (87,150,323). The enzymatic function of this protein has proved to be elusive. Only the ALSV IN protein has been shown to have any enzymatic activity, an endonuclease activity that cuts double-strand circular DNA with an LTR-LTR-like joint very near the expected site (see below) (57,97, 143,144). The relationship of this activity to the integration reaction remains to be clarified. All IN proteins exhibit nonspecific DNA-binding activity (215,366); the MLV protein has recently been found to bind specifically to the *att* site at the end of linear viral DNA (H. Varmus, *personal communication*).

The PR Protein

The PR protein is responsible for most or all of the cleavages that separate the gag and pol proteins from one another (including PR). Although its gene is located in the same place—between *gag* and *pol*—in all retrovirus genomes, the variable orientation of the reading frames in this region causes it to be expressed quite differently in the various groups. In ALSV, for example, it expressed as part of gag, and it therefore expressed much more abundantly than in other viruses. In MLV and HIV, pro is in the pol frame; in MMTV as well as in BLV and HTLV, it is in a frame distinct from both gag and pol.

PR shares sequence features and sensitivity to inhibitors with a large class of aspartic proteases, including a highly conserved sequence (Leu-Leu/Val-Asp-Thr-Gly-Ala-Asp-Lys) around the active site (268,269,384). This sequence is found only once in PR, as compared to twice in other aspartic proteases, raising the suggestion that it might be a dimer in its active form (203,271). This supposition has been confirmed, as has the relationship to other aspartic proteases, by development of a three-dimensional structural model for both the ALSV and the HIV enzymes based on X-ray crystallographic analysis (279,297,384,457).

The amino acid sequences cleaved by PR are quite variable, but they have some common features; these include a tendency toward hydrophobic residues on either side of the cleavage site, with a slight preference toward Tyr or Phe preceding the cut and toward Pro following it (327). A number of small synthetic peptides containing identified cleavage sites has been shown to provide suitable substrates for the cleavage activity (216). The apparent lack of specificity seems inconsistent with the precise nature of the cuts made in the natural precursors. Indeed, even quite sizable foreign proteins are completely resistant to cleavage by PR (384), at least in the native conformation. These results suggest that the specific sites of cleavage are imposed as much by other structural conditions as by recognition of primary sequence.

Other Enzymatic Activities

Over the years, sensitive assays have been used to detect a number of other functions, particularly those of relevance to DNA synthesis, such as DNA ligase, in retrovirus virions (86). These additional activities have not been associated with any known virus-coded protein, and it seems likely that they are either adventitiously included in virions or are included within a small fraction of cellular debris that inevitably copurifies with virions.

The Envelope

As with most enveloped viruses, the retroviral envelope is derived by budding from the cell membrane, and it comprises preexisting lipids and other membrane components modified by the insertion of viral proteins in place of normal cellular components. The composition of the membrane reflects that of the host, and there is no evidence for any virus-induced modifications to the lipids or other small molecules.

Envelope Proteins

All retroviruses have a similar complement of envelope proteins, illustrated schematically in Fig. 7. The complex is composed of two different polypeptide chains held together by disulfide bonds as well as noncovalent interactions. This association between the subunits is often somewhat tenuous, and loss of SU proteins (like HIV gp120) from virions during purification is often seen. Although this complex does not display any of the activities (such as hemagglutinin, fusion activity, or neuraminidase) seen with other enveloped viruses, its size and organization closely resemble those of the much-better-studied HA protein of influenza virus (see Chapters 4 and 21). The three-dimensional structure of a retroviral env protein has yet to be determined. Until such time as this information is available, the influenza model will provide a useful framework on which to pin an overall concept of the env protein.

The larger (SU) polypeptide contains the receptor-binding function and is also the primary [although not

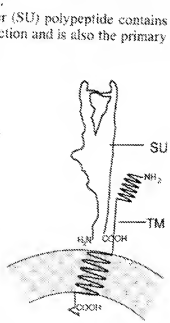

FIG. 7. The envelope protein. This diagram indicates the inferred relationship between the TM and SU proteins and the virus envelope. The helical regions denote hydrophobic amino acid sequences involved in membrane attachment and fusion.

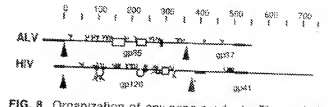

FIG. 8. Organization of *env* gene products. The scale is in amino acids. Thick bars depict hydrophobic regions. From left to right: the signal peptide; the hydrophobic "fusion" sequence; the membrane anchor region. Arrows show sites of cleavage by cellular enzymes. The branched structures depict the predicted location of N-linked polysaccharide chains. The loops in the HIV structure show the location of sequences which exhibit considerable variability from one isolate to the next (see Chapter 28), and the boxes represent regions implicated in receptor binding. (Modified from ref. 63.)

the only (302,326)] antigen capable of eliciting synthesis of neutralizing antibodies in the infected host. This protein is invariably glycosylated, sometimes quite heavily. The ALSV SU protein is predicted to have 14 sites of the sequence Asn-x-Ser/Thr which serve as addition points for N-linked carbohydrate side chains. Careful analysis suggests that the actual number of side chains corresponds with the prediction (179), implying that all predicted sites are used. The contribution of carbohydrate to the mass of the protein is considerable; the ALSV protein alone has a predicted molecular weight of around 38,000 daltons; its mobility in polyacrylamide gels is consistent with a molecular weight of about 85,000 daltons. The situation in lentiviruses is even more extreme; HIV SU proteins have over 30 such predicted sites for glycosylation (294). The distribution of glycosylation sites and other features on the ALSV and HIV env proteins is shown in Fig. 8. In addition to the N-linked carbohydrate, the presence of O-linked polysaccharide chains has also been suggested in some virus groups such as MLV and FeLV (335), but little is known about them.

In some virus groups, it has been possible to localize regions of the SU protein important for receptor binding. In ALSV, this has been accomplished by taking advantage of (a) the polymorphism of cellular receptors and (b) the corresponding diversity of receptor-binding ability among closely related subgroups of viruses (see below). Genetic analysis of env proteins from six subgroups implicates the two separated regions in approximately the middle of the protein shown in Fig. 8. This analysis also implies that these regions work together somehow, perhaps by folding to form a single receptor-binding region (37,38,90,91). In HIV, deletion-mutant analysis and competitive antibody-binding studies (218,229,464) implicate the rather differently placed region shown in Fig. 8.

The smaller env protein is the transmembrane (TM) protein. In most virus groups, with the exception of

the mammalian C-type viruses, the TM protein is also glycosylated, although usually much less extensively than its partner. As its name implies, this protein has three domains. The external region is attached to the SU protein and also contains at its amino terminus the hydrophobic region believed responsible for fusion with the cell membrane (126). The membrane-spanning domain is made up of a string of 20–25 exclusively hydrophobic amino acids (except in HIV, where it is interrupted by a lysine residue). The cytoplasmic domain is usually quite short (about 20–30 amino acids), except in some lentiviruses, such as HIV, where it is considerably longer. In MLV and related viruses, about 16 amino acids are removed from the C-terminus during virion assembly; this modification has not been described in other groups. The function of the cytoplasmic domain has not been firmly established. In some lentiviruses, such as SIV, the absence of a portion of this region does not inhibit replication in cell culture (167).

The obvious conclusion that a specific interaction between the intracellular region and a gag protein facilitates incorporation of the env protein onto the virion surface has not fared well in the face of ALSV mutants that encode protein lacking the cytoplasmic region (329). Such viruses not only have normal levels of virus production and infectivity but also have a normal complement of env protein, leaving open the source of the specific interaction responsible for ensuring the association of envelope and capsid.

Until recently, the organization of the env protein into higher-order structures had not been approached experimentally. There is now good evidence, based on sedimentation analysis, that the ALSV protein at least, and probably all others, exists as a trimer on the surface of the virion (99). This is another point of similarity between the structure of retroviruses and that of influenza and other enveloped viruses.

THE RETROVIRUS REPLICATION CYCLE

The replication cycle followed by all retroviruses is schematically outlined in Fig. 9. In brief, it consists of the following steps:

1. Attachment of the virion to a specific cell-surface receptor.
2. Penetration of the virion core into the cell.
3. Reverse transcription within the core structure to copy the genome RNA into DNA.
4. Transit of the DNA, still associated with the incoming virion proteins, to the nucleus.
5. Integration of the viral DNA into more or less random sites in cell DNA to form the provirus.
6. Synthesis of viral RNA by cellular RNA polymerase II using the integrated provirus as a template.
7. Processing of the transcripts to genome and mRNAs.
8. Synthesis of virion proteins.
9. Assembly and budding of virions.
10. Processing of capsid proteins.

In the detailed discussion that follows, there are several key points that should be kept in mind. First, the cycle has two obvious phases, divided by the integration step. The first phase is carried out by incoming virion systems that remain in a structure derived from the virion core, and the second phase is accomplished by cell machinery. Second, the provirus, once integrated, is stable. It is replicated and passed to progeny cells as part of the chromosomal complement of DNA. There is no evidence for any mechanism for removal of a provirus from the cell DNA, for its direct transposition to another site, or for its independent replication. Third, with many retroviruses, expression of the provirus does not require the help of any virus gene products. Finally, in most cases, the replication process proceeds without significant effect to the infected cell, which often continues dividing and is unaltered in any obvious way except that it is now continually producing new virions. A corollary of this relatively benign association is that the process of replication must be under some sort of control that prevents an infected cell from being repeatedly reinfected by the virions it produces, a series of events that would inevitably be fatal to the infected cell.

Attachment

Virion-Receptor Interactions

Like all viruses, retroviruses have a specific requirement for interaction with a cell-surface receptor molecule for infection. In all cases known (and suspected), this molecule is a protein that interacts specifically with the SU protein on the virion envelope. In general, it has proved much more fruitful to probe this interaction by studying the requirements for infectivity rather than by studying the binding of virions to cells. Binding of retrovirus virions to cells is often quite nonspecific when measured by simple physical techniques. Nevertheless, initiation of a replication cycle has a highly specific requirement for interaction of the virion env protein with a cell-surface receptor (459). In the absence of the appropriate receptor, infectivity can be reduced by seven or more orders of magnitude. That such a reduction is due to a block in an early event is shown by the use of cell-fusion techniques to artificially fuse virus and cell membranes (partially by-

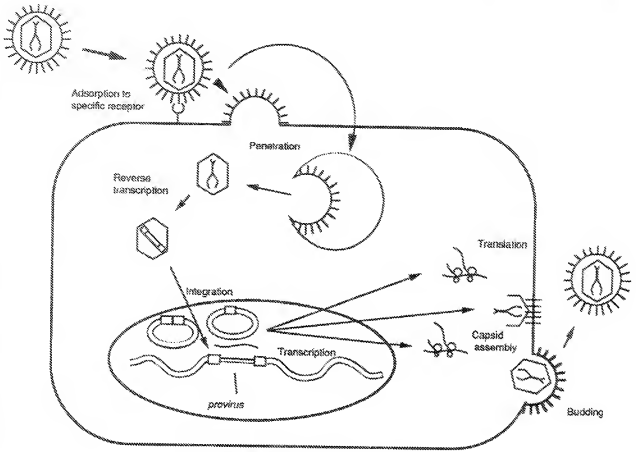

FIG. 9. Overview of retrovirus replication.

passing the need for a specific receptor) and by the use of "pseudotype" virions derived from vesicular stomatitis virus (with a retrovirus env protein substituting for the usual G protein). These display the same receptor requirements for infection as do the retrovirus env protein donors, ruling out any effects on events much later than the initial virus–cell interaction (459). The system has been particularly useful for studying virus–receptor pseudotype interactions—for example, in identifying the HIV receptor (79).

Very little is known about how retrovirus proteins interact with their receptors, and it is difficult to generalize from what is known.

By far, the best-studied interaction is that of HIV with the CD4 cell-surface protein that serves as its receptor. Although HIV is one of the most recently discovered viruses, CD4 was the first retrovirus receptor to be identified (79,213,257). This interaction will be considered in more detail in Chapter 28. In brief, the binding site on the SU protein (gp120), tentatively identified as the region shown in Fig. 8, interacts with a fairly small region on the receptor identified by the use of (a) site-directed mutagenesis and (b) recombinants with the mouse CD4 protein (which is closely related to the human one but does not bind HIV) (56,191,227,281,347). The region so identified is not in a part of CD4 directly involved in cell–cell recognition or any other known function of this protein. This binding per se seems insufficient to initiate the replication cycle, since mouse cells expressing human CD4 bind HIV but are not infected by it (257).

The only other retrovirus receptor to be identified so far is that of Moloney murine leukemia virus, which was discovered recently by use of an elaborate cloning strategy in which nonpermissive human cells transfected with mouse DNA became receptor-positive and could be selected by infection with virus containing a selectable drug-resistance gene (3a). The receptor gene could then be (a) identified by the presence of the mouse repeated DNA, (b) cloned, and (c) analyzed in detail. The protein predicted by its sequence is very different from CD4. Whereas CD4 seems to have a structure typical of a cell-surface recognition protein, with a single membrane-spanning region and a large extracellular domain (259), the MLV receptor appears to have multiple intramembrane domains and very lit-

de extracellular amino acid sequence. In this respect, its structure is more closely reminiscent of a membrane pore or channel protein. The site of interaction between the receptor and the SU protein is not known.

Genetics and Distribution of Receptors

Retroviruses are highly unusual among animal viruses in the polymorphism of receptor utilization displayed by otherwise very closely related viruses. In the absence of direct isolation, receptors can be identified and studied in one of several ways (459).

Interference

Cells infected by retroviruses display a very strong resistance to superinfection by viruses that utilize the same receptor as the preinfecting virus; moreover, these cells display complete susceptibility to viruses that use a different receptor. This is easily assessed if the second virus encodes a dominant marker such as an oncogene. This phenomenon obviously arises from interaction between the env protein of the preinfecting virus and the receptor, but the level of this interaction is not known. Virion formation is not necessary for interference, indicating that the env protein and receptor can interact in *cis* in the same membrane. The interference is virtually absolute; titers of superinfecting viruses can be reduced by more than 10^7-fold (459), suggesting that more than simple competition may be involved. In the case of HIV infection, there is a striking loss of receptor protein from the surface of chronically infected cells (175), suggesting that the interaction may occur during synthesis and/or processing of the two proteins, leading to failure of the receptor to be correctly processed.

A related phenomenon, called "early interference," can be demonstrated by preincubating cells with virions, disrupted virions, or env protein (459). This effect requires large amounts of protein and is not as strong as superinfection resistance, and it is clearly due to interaction of env protein with receptor on the cell surface.

Species Distribution

Use of different receptors can result in a distinctive host range that reflects the presence or absence of a receptor in animals of different species. Murine C-type viruses can be distinguished as follows: *Ecotropic* viruses, like Moloney MLV, use a receptor found predominantly on mouse cells, the one described above. *Xenotropic* viruses utilize a receptor which is found on many non-mouse species but which is seemingly specifically lacking in most, but not all, mice. *Amphotropic* viruses recognize receptors found on both mouse and non-mouse species, as do *polytropic* viruses. Assignment of the latter two to different groups is based on their failure to interfere with one another (343). The xenotropic and polytropic host ranges are found only in endogenous viruses of mice—viruses that have established proviruses in the mouse germline and are passed from parent to progeny as Mendelian genes (60,411).

The assignment of subgroup by chance distribution of the receptor among different species, although firmly entrenched among specialists in this area, is one of a number of facets of retroviral nomenclature that has generated considerable confusion in the field.

Polymorphism Within a Species

In chickens, individual animals (or lines of animals) of varying sensitivity to infection with different ALSV isolates are commonly found (459). Based on this variation in susceptibility, as well as interference patterns, no less than six distinct subgroups, denoted A through G, have been distinguished. (One subgroup, D, is a variant of another, B, and uses the same receptor but differs slightly in other properties.) Subgroups A through D are found among exogenous viruses of chickens; E is unique to endogenous viruses of chickens, and F and G are found among endogenous viruses of pheasants. The ability of a cell to resist infection by a specific subgroup is denoted by a slash (or bar) following a letter standing for the species and followed by the subgroups to which the cell is resistant. Thus, C/A ("C bar A") are chicken cells sensitive to all virus subgroups except A; T/BD are turkey cells resistant to B and D subgroups; and C/O ("C bar oh") cells are susceptible to all known subgroups.

Genetic analysis has revealed the existence of three unlinked autosomal loci responsible for these interference patterns (459). These are known as *tv-a*, *tv-b*, and *tv-c*. In all cases, sensitivity is dominant to resistance, implying that the loci encode the receptors themselves, although more complex explanations are still possible. Loci *tv-a* and *tv-c* have two alleles, encoding sensitivity and resistance. Locus *tv-b*, which controls sensitivity to both B and E, is more complex; three alleles are known, encoding resistance to both, sensitivity to both, and resistance to subgroup E (the most common pattern), respectively. The fourth combination, sensitivity to subgroup E virus and resistance to subgroup B virus, is not seen in chickens but is the most common pattern in other birds, such as turkeys and quail.

The inability of many endogenous viruses to infect the animal whose germline they inhabit is not uncommon. The polymorphism among receptors for retro-

viruses in a given species may reflect genetic changes in the species in response to selective pressure for resistance to infection with the virus sometime in the past. It is important to bear in mind that the lack of receptors in a given species does not necessarily reflect lack of the protein; instead, it probably reflects sequence differences in the protein itself from one species to another. The use of such differences can be of great help in localizing the specific binding domains on a receptor molecule, as in the case of CD4 mentioned above.

Penetration and Uncoating

The mechanism by which retroviruses enter cells is one of the most poorly understood aspects of the virus life cycle. In the absence of a good structural model for the env glycoprotein, and in the face of a very poor ratio of infectious to total virions (meaning that the vast majority of observable events are nonproductive ones), only indirect and inferential evidence as to the mechanism involved is available. Comparison with the influenza model suggests the following sequence of events: After binding of the SU protein to its receptor, the virus envelope and the cell membrane fuse to release the virion core into the cytoplasm. This fusion can take place at one of two sites. In the case of HIV, this fusion probably occurs at the cell surface immediately after binding. This conclusion is suggested by the tendency of cells expressing HIV env protein to fuse directly to uninfected, receptor-positive cells (240,388) (see Chapter 28). Furthermore, penetration of HIV occurs at a pH typical of the cell surface, not at the relatively acidic conditions that are present following endocytosis (401); and CD4 mutants, themselves incapable of endocytosis, nonetheless serve as receptors (18,258). Most other retroviruses do not cause cell fusion, and infection by at least some is sensitive to high pH (336). Like many enveloped viruses, they seem to be internalized via receptor-mediated endocytosis followed by fusion of viral envelope and endosomal membrane, possibly provoked by the lower pH typical of the endosomal contents.

Little is known about how the fusion itself occurs. It seems likely that this step is mediated by the region of hydrophobic amino acids at the amino terminus of the smaller TM protein. Site-directed mutagenesis of this region in HIV and other viruses has yielded results consistent with this idea (270,330). Mutational alterations that prevent cleavage of the two proteins to generate the hydrophobic end, as well as mutations that insert charged amino acids into it, preserve receptor binding but block infectivity and cell-fusion ability of the virus. The important interaction of the fusion domain could either be with the membrane itself or with some component (such as a protein) within it. The presence of a specific "fusion receptor" for HIV has been invoked to explain the lack of infectibility of mouse cells expressing human CD4 protein (257), as well as to explain the sequence similarity between (a) the hydrophobic region in HIV and (b) the F protein of some paramyxoviruses that infect human cells (126).

The fate of the various capsid proteins upon fusion is not well known. Clearly, RT, IN, and NC proteins must remain with the genome, and there is good reason to believe that CA does as well (see below). It seems not unreasonable to suppose that the MA protein remains associated with the membrane, but there is no evidence regarding this point.

Synthesis of Viral DNA

Overview

After entry of the core into the cytoplasm, the process of reverse transcription of the RNA genome into double-strand DNA occurs, using the structure and enzymatic activities that entered the cell in the virion (442,446,447). An outline of the process of synthesis of viral DNA is shown in Fig. 10. The mechanism by which this occurs is a complex but elegant solution to two problems that a virus replicating by this mode would inevitably face. First, the requirement of the system for RNA primers precludes precise end-to-end copying of the genome into DNA, since there would be no way to copy the primer-binding region. Second, since new viral genomes are to be made by cellular RNA polymerase II, some way must be provided to ensure that signals appropriate to direct the synthesis of RNA lie upstream of the initiation site of synthesis, outside of the region to be copied. Obviously, such signals cannot simultaneously be inside and outside of the viral genome. The resolution of this paradox evolved by retroviruses and related elements is to provide for the synthesis at each end of the DNA molecule of an extra copy of sequences present only once in the RNA genome. These extra sequences together form a structure—the long terminal repeat (LTR)—which contains virtually all of the *cis*-acting sequences necessary for events that take place at the DNA level, namely, integration and expression of the provirus.

Only one enzymatically active protein molecule—the reverse transcriptase—seems to be necessary to accomplish this unusual metamorphosis, although completely correct synthesis also requires the structural milieu provided by the capsid proteins. The secret of viral DNA synthesis is in the "jumps"—the ability of the enzyme, on encountering a block to DNA synthesis, to transfer the growing DNA chain to a similar sequence elsewhere and then continue elongation.

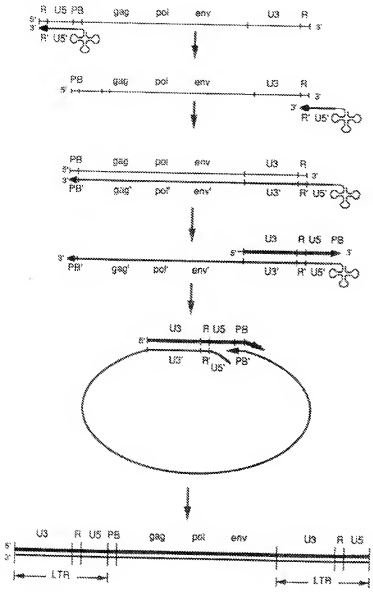

FIG. 10. Mechanism of viral DNA synthesis. Thin lines depict RNA, medium lines depict negative-strand DNA, and thick lines denote positive-strand DNA. Negative-sense sequences are also indicated by a "prime." (Drawing courtesy of S. Herman.)

Two such transfers form the LTR by moving copies of the U5 and U3 regions flanking R to either end of the viral DNA. Specific sequences at the tips of these regions then provide signals for integration.

The In Vitro *Reaction*

The first inkling that retroviruses replicated through a DNA intermediate—a heretical idea when first proposed and for many years afterward (428)—was provided by experiments on the ability of certain inhibitors of DNA synthesis to block early steps in replication or on the ability of DNA-directed RNA synthesis to block late steps. This idea received dramatic confirmation with the demonstration that retrovirus virions contain an enzymatic activity that can copy the virion RNA into DNA (12,429). This activity can readily be seen when virions, rendered permeable with a nonionic detergent, are incubated with an appropriate buffer containing deoxynucleoside triphosphates, one of which is usually radioactively labeled. Synthesis of DNA complementary to at least a fraction of the genome then ensues. Under carefully controlled conditions, a full-length, double-strand, biologically active copy of the genome, complete with LTRs, is synthesized, validating the use of the *in vitro* reaction to study the mechanism. A useful adjunct to this reaction is the addition of actinomycin D to block DNA-directed (positive-strand) synthesis but not RNA-directed (negative-strand) synthesis. In this way, it has been possible to obtain a very detailed picture of the overall process.

Synthesis of Negative-Strand DNA

If the *in vitro* reaction products are separated by their size, a number of discrete species, representing points at which reverse transcriptase pauses during elongation, can be seen. The earliest products are complementary to the genome RNA and are still attached to the tRNA primer at their 5′ ends. The most prominent of such products is a molecule called "strong-stop" DNA, which is a copy of the short region consisting of R and U5 lying between the primer-binding site and the 5′ end of the genome (65). Its presence as a major species *in vitro* implies that DNA synthesis initiates by elongation from the 3′ end of the primer tRNA until the 5′ end of the genome is reached.

Once strong-stop DNA is made, the newly made strand must transfer, presumably along with the reverse transcriptase, to the other end of the genome in order for synthesis to continue. Several features of the system make this transfer, or "jump," possible. First, the R sequence, present near each end of the genome, permits appropriate base-pairing to direct continued synthesis to the correct position. Second, the RNase H activity of reverse transcriptase removes the newly copied RNA, leaving the DNA free to pair with the R sequence at the other end. Consistent with this requirement, mutations in *pol* which specifically affect RNase H activity strongly inhibit this transfer (422; S. Goff, *personal communication*). Third, some unidentified aspect of the capsid structure is necessary for this process. Although all the synthetic and nucleolytic activities of reverse transcriptase can be duplicated in reactions with purified enzyme and template, the jump has never been accomplished in pure reconstituted systems. Indeed, the synthesis of full-length negative-strand DNA, even by disrupted virions, drops off dramatically at concentrations of detergent even slightly greater than optimal. Instead, the major product at greater-than-optimal detergent is strong-stop DNA along with a significant fraction of incorrectly elongated molecules, as if the jump had been to the wrong part of the genome. It appears that the driving force for jumping is more likely to be affinity of reverse transcriptase for its template; it also appears that the major responsibility for ensuring correct jumping lies in capsid structure, not in sequence homology. The complementarity between the R region at the 3′ end of the genome and its copy in the DNA serves to align synthesis properly but does not itself direct the jumping. Indeed, experiments designed to test the heritability of mutations in either copy of R suggest that transfer can occur prior to copying its full length (248).

A final suggestion regarding the importance of capsid structure to reverse transcription is raised by another experiment. The contributions to the two ends of the provirus of sequences derived from the ends of either parent was assessed in clones of progeny infected with virions heterozygous for markers at either end (320). The strong tendency of the proviruses to contain U5 sequences from one parent and U3 from the other implies that the jump may be directed specifically to the 3′ end of the genome molecule rather than to the end at which initiation occurred. This result is difficult to understand, since the organization of RNA in the capsid must be very complex; furthermore, it is far from obvious how the ends could be tagged so that they can be distinguished by the system, and there is no compelling logical requirement for the use of two different templates at this or any point in the process.

Once the jump has occurred, synthesis of the negative strand can proceed unchecked to the 5′ end of the template, which is now the 5′ end of PB because R and U5 will have been removed earlier by RNase H. It is possible that this process is not continuous but, rather, might be frequently interrupted as reverse transcriptase encounters breaks in the RNA template. In this case, jumps analogous to those occurring at the end of the genome may occur, leading to recombination and repair of genomic damage (see below).

The molecule that results from copying the genome is a complete, but slightly permuted, complement of the genome of the structure

3′—PB–gag–pol–env–PB–U3–R–U5—(tRNA)—5′

The tRNA primer is still attached at the 5′ end and is in a DNA–RNA hybrid that is base-paired with the genome.

Synthesis of Positive-Strand DNA

To obtain a completed molecule, it is necessary for reverse transcriptase to copy the negative-strand DNA molecule just made. This process also requires a primer molecule to initiate synthesis. In this case, the 3′ end that serves this role does not preexist; it must be created. This is accomplished by a specific cleavage of the RNA template at the 5′ end of U3. This specific cleavage seems to occur by a variation of the RNase H reaction; it requires that the RNA sequence to be cut be in a hybrid configuration. It can be readily duplicated *in vitro* using highly purified reverse transcriptase and model substrates (342). Unlike the usual reaction catalyzed by RNase H, cleavage at this site does not seem to require the proximity of an end and shows great specificity. All molecules are cleaved to yield an end at the same location, since close observation of the DNA molecules primed at this site both *in vitro* (113) and *in vivo* (125) reveals no detectable

heterogeneity. The signal for this specific cleavage lies in the polypurine (PP) tract characteristic of all retroviruses and retrotransposons (113), although what part of this sequence forms the signal itself remains to be worked out.

Although cleavage adjacent to the PP tract is quite precise and occurs with every molecule, it is possible to detect initiation of positive-strand synthesis at other points on the genome as well, both *in vitro* and in the infected cell. At least some of these are probably the result of the same kind of reaction; the sequence upstream of these sites resembles the PP region (113). In other cases, the preexisting breaks in the genome mentioned above could serve as priming sites after negative-strand synthesis. The extent of use of such false priming sites varies from one virus group to another; analysis of DNA made in infected cells shows a much greater proportion of fragmented positive strands in ALSV than in MLV infections. Note that although the synthesis of the positive strand is usually shown as a continuous process from end to end (as in Fig. 10), positive-strand DNA could also be made in small pieces and subsequently joined by cellular ligases, provided that initiation was also at the correct site on the same molecule.

Failure to use the correct site for initiation of positive-strand synthesis can result in the generation of viral DNA forms containing either more or less than a full LTR. Such forms are frequently seen in cloned viral DNA molecules (312,378,444). Since the lack of a proper LTR tip will block integration, such molecules are likely to be overrepresented among the unintegrated DNA molecules usually used for cloning.

Following initiation, elongation of the positive strands is carried out by reverse transcriptase toward the 5′ end of the negative strand, which is soon reached. As with the negative strand, another jump must occur to permit complete synthesis. In this case the redundant sequence used is formed by the copy of the primer-binding site at the 5′ end of the negative strand. As can be seen in Fig. 10, the positive strand is elongated by copying the negative strand DNA through the U3, R, and U5 regions and into the primer itself. Copying the first 18 bases of the primer yields a direct copy of the primer-binding site which can then form yet another template–primer pair with its complement at the 3′ end of the negative strand (133). The species of positive-strand RNA just prior to the jump is a prominent product of reactions *in vitro* and *in vivo* (446), and it is called "positive-strand, strong-stop" DNA. The termination point at the 3′ end of positive-strand, strong-stop DNA is set not by the end of the tRNA primer but, rather, by the presence of a modified base in the tRNA molecule itself, an m^1A residue that cannot be copied. Termination at this point ensures a perfect match of the 3′ end of positive-strand strong-stop DNA with the PB sequence at the end of the negative-strand DNA.

Completion of the full-length, double-strand DNA can now be accomplished by the synthesis of each strand to its logical end. All that is required is to clean up some details. The primer tRNA must be removed from the 5′ end of the negative strand. This reaction presumably uses a special feature of the RNase H activity. It can be readily duplicated with purified reverse transcriptase and artificial substrates (313). The genome RNA is most likely completely degraded by the action of RNase H to make way for the synthesis of positive-strand DNA. This process makes retroviruses one of the few (if indeed there are any other) examples of destructive replication, in which there is no net gain of genome-related molecules after the first events in replication. Rather than carrying out any genome replication, reverse transcriptase catalyzes a metamorphosis of the genome from two molecules of single-strand RNA to double-strand DNA. The number of DNA molecules—one or two—is not known with any confidence. There are some less-than-compelling reasons to believe that there might be only one.

Sites and Structures for DNA Synthesis

A number of lines of evidence give strong support to the idea that early events in retrovirus replication take place not in a more or less free solution (as they are usually depicted) but, rather, within a structure derived from the viral capsid. First, proper completion of viral DNA synthesis *in vitro* seems to be critically dependent on the capsid structure. Second, there is a genetic locus in mice (called *Fv-1*) that encodes intracellular resistance to infection by some MLV strains. The block to infection seems to occur following viral DNA synthesis but prior to integration. Virus mutations that relieve the block have been found to map to the region encoding the CA protein (34,317,357), implying that this protein is still in association with the virus at the end of viral DNA synthesis and possibly later. Third, association between newly made viral DNA and the CA protein has recently been found for both MLV (39) and ALSV (D. Schenkein, *personal communication*) and is well established for ty1 (30,129) and other reverse transcribing systems (123). In both MLV and ALSV, CA protein derived from the viral capsid and unintegrated DNA cosediment in a structure much larger than the size expected for DNA alone. The MLV structure has also been shown to be active in the *in vitro* integration reaction (see below) (45,125). These structures are poorly understood. In addition to CA and DNA, it is reasonable to suppose (but is not directly shown) that they also contain RT and IN proteins, and possibly NC as well. The structures are not

very tight, and they do not protect the DNA from digestion with nucleases (39).

The site of viral DNA synthesis also remains to be well clarified. Cell fractionation using detergent treatment implied that DNA synthesis takes place in a soluble cytoplasmic fraction (446). More recent experiments suggest a detergent-soluble compartment associated with a large structure (D. Schenkein, *personal communication*). Whether the signals (as well as subsequent translocation to the site of integration) reside in the DNA, the CA protein, or elsewhere remains to be determined, as does the mechanism of entry of the structure into the nucleus.

Following DNA synthesis, which takes place rather asynchronously in the first 4–8 hr after infection, the viral DNA becomes tightly associated with or within the nucleus. Two additional forms appear: covalently closed circles containing either one or two LTRs. The circles containing one LTR could be formed from linear molecules by homologous recombination across the LTRs, but it is more probable that they represent an aberration of the DNA synthetic process in which the 5′ end of the positive strand is not displaced from the negative strand at the final step (see Fig. 10). The two-LTR circles include a variety of structures derived either by ligation of the ends of the linear molecule or by integration of the viral DNA into itself (see below). Once thought to be intermediates in integration (322), recent evidence now strongly suggests that they are side products and functional dead ends (45,125). The circles have been quite useful, since they are much easier to clone than the integrated or linear forms.

Integration

Integration is the process most unique to retrovirus replication. Other animal and plant viruses use reverse transcription as part of their replication cycle (see Chapter 38); no other has a regular mechanism for stably associating itself with the host DNA. With possibly a few very special exceptions (153) (see Reference 478), integration of viral DNA is a key part of every replication cycle. The large majority of the viral DNA molecules synthesized are integrated during every round of replication, and integration is probably necessary for proper expression of the provirus as well. These properties clearly distinguish retroviral integration from the sort of "integration" that is an occasional aberration of some DNA virus infections (Chapter 13). In other systems, the closest analogy is found in bacteriophage mu (74), but there are considerable differences. For a long time, integration was the least well understood segment of the virus replication cycle, accessible only through (a) analysis of the structures that result and (b) a limited amount of genetic and biochemical information. Recent advances in the development of *in vitro* systems that accurately reproduce the process have significantly increased the depth of our understanding in the last few years.

Structure of Integrated Proviruses

Nucleotide sequence analysis of a large number of integrated proviruses has revealed a collection of features common among all virus groups, including retrotransposons (Fig. 11) (30).

First, the provirus is colinear with the viral DNA as synthesized by reverse transcriptase and contains the genes in the order they are found in the genome, flanked by LTRs.

Second, both viral and cellular DNA have undergone characteristic changes as a result of the integration process. The viral DNA has been shortened, usually by the loss of two bases (usually AA) at each end. [HIV is an exception in that only one base pair is removed from each end (294,437).] The cell DNA flanking the site of integration has not been grossly perturbed. A short sequence (4–6 base pairs) at the cell target immediately adjoining the viral DNA is always duplicated. The length of the duplication is characteristic of the virus group, not of the cell type, showing that specificity for the duplication (most likely created by repair of a staggered cut in the target DNA; see below) resides in the virus.

Third, the ends of the viral DNA are always 5′ TG.....CA 3′. This dinucleotide is found at the ends of virtually all retroviruses, retrotransposable elements, and even bacterial transposable elements. This striking conservation would seem to imply that these specific bases play some indispensable role in the process. It is thus surprising that at least one of them (the A) can be changed to a G without ill effect on MLV replication (68).

Finally, the ends of the LTR are usually characterized by an approximate inverted repeat spanning 2–10 bases. These might associate to form some special secondary structure, but it is more likely that they contain a specific recognition signal for the integration machinery. In general, mutations introduced into these sequences have a substantial damping effect on the integration process (57,68,100,321).

Specificity of the Integration Reaction

On the part of the virus, integration is a highly specific process, always making the joint to cell DNA at the same place in the viral sequence. Analysis of a series of mutations introduced into MLV genomes near the ends of the LTRs suggests that this specificity is provided both by proximity of the sequence to the ends

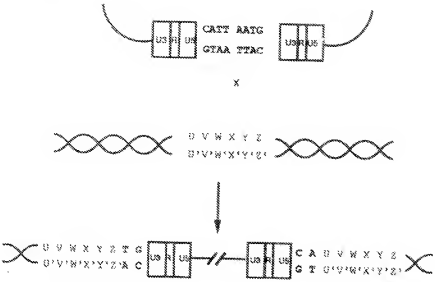

FIG. 11. Properties of retrovirus integration. The top portion shows the juxtaposed ends of unintegrated ALSV DNA and target cell DNA, and the bottom portion shows the structure after integration. Note the loss of two base pairs from each end of the provirus, and also note the reduplication of six (in this case) base pairs of cell DNA.

of the viral DNA and by recognition of a specific sequence at each end containing the inverted repeat (68; S. Goff, *personal communication*). This sequence, sometimes called "*att*," probably forms the only signal required in *cis* by the viral integration machinery. Mutational analysis has narrowed the sequence required to about 8 or 12 bases from the ends of the LTR in at least one virus (321).

From the perspective of the cellular target, most approaches to the problem suggest that integration is a much more random process. Comparison of cellular sequences flanking a modest number of independent integration events (12–20 in most cases) has revealed no common sequence which might serve as a cellular target (444,446). Similarly, analysis of the distribution of integrations into target DNA *in vitro* is not suggestive of any specificity for a particular site. Another approach to the same issue is to compare sites of integration selected for insertion into the same general region (e.g., in tumors induced by activation of protooncogenes). In B-cell lymphomas induced by ALV inoculation of chickens (see below), the large majority of tumors have an ALV provirus inserted within the first intron of the *c-myc* protooncogene (158). Within this region there does not seem to be a strong tendency to cluster at any one site (although some clustering can be noticed; see below), again consistent with the idea that there is no strong sequence selectivity for integration (138,352).

There are several reports that suggest a tendency for integration to occur in regions of DNA that might tend to be transcriptionally active (i.e., regions characterized by relatively "open" chromatin structure). Proviruses randomly cloned from MLV-infected cells were often found to be integrated in the vicinity of chromatin structural features marked by high sensitivity to DNase (356,448,449), and the *myc*-associated ALV proviruses mentioned above showed a tendency to be located in proximity to one of five such hypersensitive sites in *c-myc* (352). Such sites often mark actively transcribed genes, and their preferential use in at least some cases might provide a mechanism by which the provirus could ensure that it be inserted into the most favorable possible region for its own expression. Mechanistically, this might be readily accomplished if integration has a tendency to occur in regions of chromatin relatively devoid of competing protein.

A final and more dramatic level of specificity has been reported for a fraction of integrations of ALSV into avian cell DNA (376). In this instance, a selective cloning strategy was used to obtain large libraries of integrated proviruses joined to flanking DNA. Analysis of the flanking DNA revealed that some 20% of integrations are into one of about 1,000 sites. Even more surprisingly, independent integrations into the same region were at exactly the same place, implying a frequency of use about a millionfold greater than random. The basis for the high level of usage of these specific sites is unknown. It is tempting to speculate that they might represent regions of DNA to which the incoming virion structure has unique access (e.g., loops of DNA near the portal by which the structure containing the viral DNA first enters the nucleus).

Mechanism of Integration

Until recently, it had not been possible to approach the biochemistry of integration in any detail due to the lack of a suitable *in vitro* system. The key to the recent development of such a system was the realization that, like reverse transcription, integration of viral DNA occurs in the context of a specific structure derived from the viral capsid. Not surprisingly, repeated attempts

to obtain correct integration using purified components were completely unsuccessful. The more fruitful approach relies on the use of extracts obtained from infected cells at a time (about 12–24 hr after infection) when viral DNA synthesis is complete but while integration is still ongoing. If an appropriate target DNA is added to such an extract (either from nuclei or cytoplasm), then integration of the viral DNA in the extract into the target can be readily detected. Initially, a very sensitive assay was developed employing a modified virus containing a bacterial suppressor tRNA gene and also employing (as target) a bacteriophage cloning vector whose replication was absolutely dependent on the suppressor. Integration of viral DNA into the target could then be detected by the appearance of phage capable of replicating after introduction of the DNA into appropriate bacterial targets (45). This relatively cumbersome method has given way to a more direct analysis in which integration of the viral DNA into a small circular DNA target is detected by appropriate restriction enzyme digestion, gel electrophoresis, and hybridization. These approaches have been applied to analysis of MLV (44,125) and ALSV (Y. M. Lee, *personal communication*) DNA integration as well as that of the yeast retrotransposable element, yl (99). In all cases, detailed analysis of the reaction products show that the *in vitro* reaction is correct. There is a loss of the appropriate number of bases from the end of the viral DNA, and there is also a duplication of the correct number of bases of cellular sequence flanking the provirus.

Using these approaches, the following information has been gleaned regarding the integration process:

1. Concordant with the rationale for this approach, integration is mediated by a ribonucleoprotein complex derived from the viral core (39; D. P. Schenkein, *personal communication*). This complex has a sedimentation coefficient of about 160–180S. (Linear DNA alone would sediment at about 20S.) If isolated from nuclei, the complexes contain the linear as well as the two circular forms of DNA. Cytoplasmic structures often have only linear DNA; both nuclear and cytoplasmic forms are equally active in the integration reaction, implying that the linear form of DNA may be the important intermediate. The organization of these structures is unknown. The only protein conclusively associated with the structure is CA (39); IN and RT are likely to be present as well but have not been detected. This system is not particularly tight; the DNA within is readily accessible to added nucleases. Attempts to use nuclease digestion to probe its organization (as can be done with chromatin) have not yet been successful. The structure itself must be derived from the viral capsid: encapsidated genomes containing no viral genes whatsoever are converted to DNA and are integrated with high efficiency. At least in the case of MLV, structures purified from soluble cellular components by sedimentation or chromatography are still active in integration, implying that no soluble cell factors are required (39,125).

2. The reaction carried out by the MLV system seems limited to integration into the added target. No side reactions, such as circle formation or autointegration, are seen (44,125). In the case of ALSV, however, the situation seems to be somewhat more complex. Circular products resulting from integration of the viral DNA into itself, as well as single LTR circles, are found among the reaction products in high yield (Fig. 12) (Y. M. Lee, *manuscript in preparation*). Presumably, there are specific mechanisms to block their efficient formation *in vivo*, although small amounts of one and two LTR circles as well as the circular products arising from integration of viral DNA molecules into themselves can be found in infected cells (379).

3. The ends of the viral DNA prior to integration have a characteristic structure. The 5′ end of each strand is precisely at the site of its initiation and has therefore been modified only by removal of the primer. The 3′ ends, by contrast, are missing two bases (44,125). The ends thus have the sequence

```
5′ AATG.....CA   3′
3′   AC.....GTAA 5′
```

The formation of the truncated 3′ end could be caused either by premature termination of synthesis or by cleavage; some kinetic evidence suggests the latter mechanism (J. Murphy and S. Goff, *personal communication*; Y. M. Lee, *personal communication*). In any case, genetic experiments implicate the IN protein. A number of mutations have been introduced into the relevant region of pol; these specifically block integration but not viral DNA synthesis (87,150,323) and also block formation of the undercut 3′ end (44). Furthermore, mutations in the sequence adjoining the ends of the LTR (the so-called "*att* sequence") either modify the site at which this cleavage occurs or prevent it altogether (J. Murphy and S. Goff, *personal communication*). In the case where the mutation modifies the site of cleavage, the *in vivo* effect is as expected; the point of integration exactly corresponds to the site of cleavage. Since the majority of unintegrated forms can have this structure some time prior to the integration reaction, its formation must not be tightly coupled kinetically to the rest of the integration reaction.

The integration reaction *in vitro* joins the previously formed 3′ end to a 5′ end of the target DNA (44,125; Y. M. Lee, *personal communication*). The 5′ ends of the viral DNA are unaltered and remain unjoined. This result strongly implicates a linear intermediate rather than a two-LTR circular intermediate as was previously thought (322). Cleavage of a circular form in

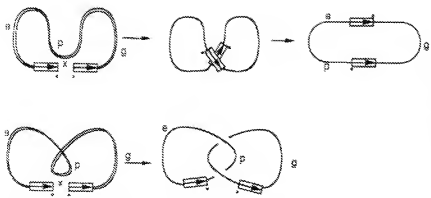

FIG. 12. Autointegration of viral DNA. The outcome differs depending on the conformation of the DNA at the time of integration. In the top case, the product is a two-LTR circle with one region (including one LTR) inverted with respect to the other; in the bottom case, it consists of two (usually, but not necessarily) linked circles, each with one LTR. (Courtesy of Y. M. Lee.)

which 5′ and 3′ ends are joined to give the 3′ structure seen would also have given a 5′ end two bases longer than observed. The joining of the 5′ ends to the target is currently thought to be accomplished by cellular repair systems.

4. There is no obvious requirement for a particular sequence or structure of the target DNA. Although some experiments discussed above imply some importance to chromatin structure, the presence of histones or other DNA-binding proteins does not seem to be necessary for the reaction to occur. The effect (if any) of the chromatin structure itself on the rate or specificity of integration has not been investigated.

5. At least in the case of MLV, the reaction proceeds efficiently in the absence of ATP or of any other added energy-generating system (45). This independence suggests that the reaction may resemble the sort of coupled breakage–rejoining reaction catalyzed by topoisomerases, rather than separate cleavage and ligation. Before this point can be proven, however, it must be shown that the system itself is not precharged energetically in some way.

All of these considerations have led to formulation of the pathway shown in Fig. 13. While reasonable and consistent with all data, a number of the steps shown should be considered as hypothetical. The overall process would involve the following steps.

1. Following viral DNA synthesis, the core structure containing linear DNA and the CA and IN proteins (and possibly the RT and NC proteins) enters the nucleus.
2. The 3′-terminal two bases at either end are removed by the IN protein, leaving a 3′ OH end. This reaction may occur before entry into the nucleus.
3. A concerted reaction introduces a staggered cut into a random site of the target with 4–6 bases (depending on the virus) of overhanging 5′ phosphorylated end and joins it to the 3′ ends of the viral DNA.
4. A cellular DNA repair system fills in the resulting gap in the molecule, displacing the two mismatched bases at the 5′ end of the provirus and ligating the remaining ends. This gap repair of the initial staggered cut generates the characteristic duplication of cell DNA flanking the provirus.

The Role of Viral Proteins

The enzymatic activities responsible for most of the steps have not been firmly identified. Although the IN protein is generally believed to carry out the major steps, genetic evidence strongly implicates it only in the first reaction—removal of the 3′ two bases from the ends of the viral DNA. Biochemical evidence is even more scanty. Although the ALSV protein has an endonuclease activity that exhibits some preference for cutting viral DNA near the expected site [although on the wrong molecule—only circular substrates have been tested to date (57,97,143,144,214,431)], the MLV and SNV proteins, whether purified from virions (319) or from bacteria expressing an appropriate cloned construct (256,364), exhibit no enzymatic activity, only a predilection toward nonspecific binding of DNA. Thus, the question of which protein actually performs the marriage of virus to cell DNA remains open. All known mutations in IN have been found to block the steps in replication prior to integration, making the genetic approach noninformative on this point. Other approaches, including biochemical dissection of the system to its component parts, will be required in order to answer these questions.

Stability of the Provirus

Once integrated, the provirus, for almost all purposes, can be considered to be perfectly stable. There is no specific mechanism by which proviruses can be excised, moved from one location to another, or replicated independently of the chromosome in which they reside. When occasional loss of a provirus is seen, it appears to be due to random cellular processes, not to virus-encoded functions. Such processes can include deletion of a larger region of cell DNA including the provirus, or, more commonly, recombination between the LTRs, leaving behind a single LTR. The

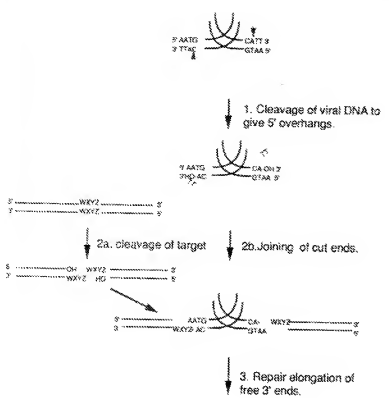

FIG. 13. A possible mechanism of integration. The scheme shown is consistent with experimental results, but all steps have not yet been proven. Note that the generation of the recessed 3′ end is temporally unlinked to the subsequent events, but also note that cleavage of the target and joining of the viral DNA may occur as a concerted reaction (45,125). The final step may be accomplished by a cellular repair system.

best-studied case is that of the endogenous provirus of mice whose insertion was responsible for the dilute (*d*) coat color mutation of mice (193). Since recombination across the LTRs causes reversion of the mutation, and large numbers of *d* mice are raised each year, it is relatively easy to screen for such revertants. When such a survey was done, a reversion rate of about $4–5 \times 10^{-6}$ per generation was estimated (70,374). Somewhat lower values (about 10^{-7}) have been estimated for the *d* mutation in somatic tissue and for proviruses in tissue culture cells (445).

Expression of the Provirus

Once the provirus is integrated into cell DNA, the virion systems have accomplished their purpose and all further replication is via transcription of the provirus into RNA using cellular systems. This process can be quite efficient, since it has been estimated that up to 10% of the mRNA in an infected cell can be derived from one or a few integrated proviruses (446; J. P. Stoye, *personal communication*). The overall strategy of expression of some retroviruses is shown in Fig. 14. The provirus is transcribed into a single RNA precursor, which is subsequently processed by (a) polyadenylation at the 3′ end of R to yield a genome-length molecule and (b) splicing of a fraction of the transcripts to generate at least one subgenomic mRNA species. After transport to the cytoplasm, a fraction of the full-length RNA is reserved for genomes, and the remainder is used as mRNA for *gag*, *pro*, and *pol*. These genes are expressed as a nested set of precursor polyproteins. The gag, gag-pro, and gag-pro-pol precursors are found in various combinations in different viruses, in graded amounts such that the pol peptides are about 5% as abundant as gag peptides. This relationship results from partial bypassing of the translational stop signal at the end of gag. The spliced RNA is translated on membrane-bound polyribosomes to give the env precursor protein. Other spliced mRNAs, if present (as they are in lentiviruses, for example), yield a variety of gene products that are not found in virions but, rather, act in varied and complex ways to regulate the expression process itself.

Transcription

A major function of the LTR is to provide signals recognized by cellular transcription machinery for the efficient expression of the provirus. Transcription of

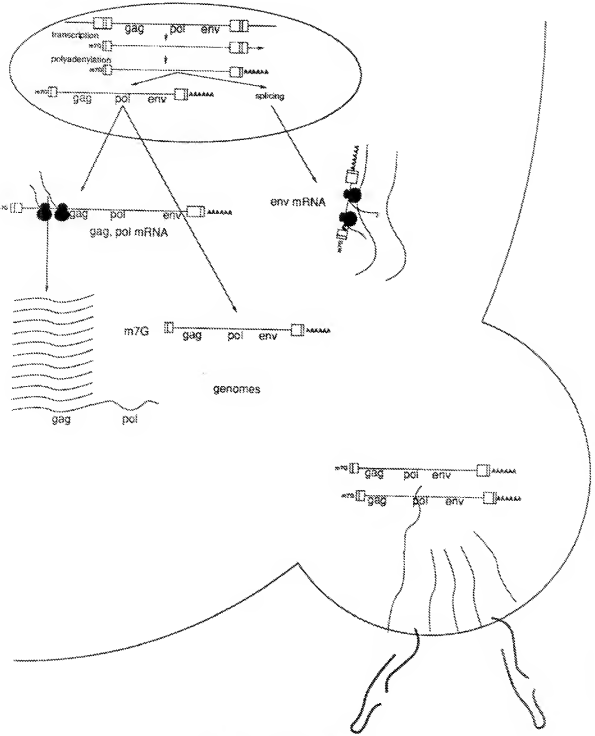

FIG. 14. The strategy of retroviral gene expression.

the provirus is initiated at the site where the capping group is placed (by definition the U3–R junction (or cap site)), and it apparently proceeds through the 3′ LTR into flanking cell DNA, with the final 3′ end determined at the end of R by cleavage and poly(A) addition. All retrovirus genomes are synthesized by RNA polymerase II, the same enzyme responsible for synthesis of cell mRNA; moreover, the LTRs of different retroviruses all contain sequences clearly identifiable with corresponding sequences in normal cellular genes, although often in complex combinations. Virtually all these sequences lie upstream of the initiation site of transcription, and thus they lie within U3. The disposition within several different LTRs of sequences

known or suspected to be important for transcription is shown in Fig. 15. Several different lines of evidence have pointed to the role of these sequences in permitting and regulating virus expression.

Within most groups of viruses, the greatest sequence divergence is found within U3 (61,62,437). Among members of the ALSV group, for example, the U3 region of endogenous viruses is quite different from that of their exogenous counterparts, despite the close similarity of the rest of the genome. Endogenous avian retroviruses, when obtained as infectious viruses, also exhibit a replication rate some 10- to 30-fold lower than that of exogenous viruses. The correlation between these differences is shown by analysis of viruses with recombinant genomes. All recombinants analyzed replicate with the rate characteristic of the U3 region donor (436). This difference in replication rate is important to the biology of these viruses. The reduced replication rate is most likely an essential feature of endogenous viruses, since it reduces their ability to induce disease in the infected host (43,351) and thus permits their benign association with the germline.

A similar example comes from the murine leukemia virus group. Again, differences in sequence affecting pathogenesis, differences in the ability to induce tumors (35,380), or differences in the specific target cell for infection and/or transformation (36,120,137,239, 358,398) are found to lie within U3. In this case, these differences may be quite small; only a few bases can determine the relative specificity. These differences lie within the region of U3 that contains the enhancer element, and, as one might expect, they seem to affect the interaction of LTR sequences with one or another cell transcription factor (396). Introduced mutations such as the insertion of regulatory regions from other LTRs, or even specific enhancer sequences from other groups of viruses, can have similar effects. For example, most MLV strains do not replicate in cells of early developmental stages, such as stem cells *in vivo* (141,465) or undifferentiated embryonal cells (ECs) in culture (112,141,425). However, when ECs are induced to differentiate, they become permissive for MLV infection, implying the appearance during development of a different panel of transcription factors. This failure has been traced to the lack of appropriate transcription signals in the LTR as well as the presence of possible negative regulators (141,246,252,253,304). At least one natural variant can provide a somewhat different U3 sequence capable of supporting expression and replication in these cells (120,164,406). Also, insertion into the U3 region of the enhancer region of a mutant strain of polyomavirus capable of growing in

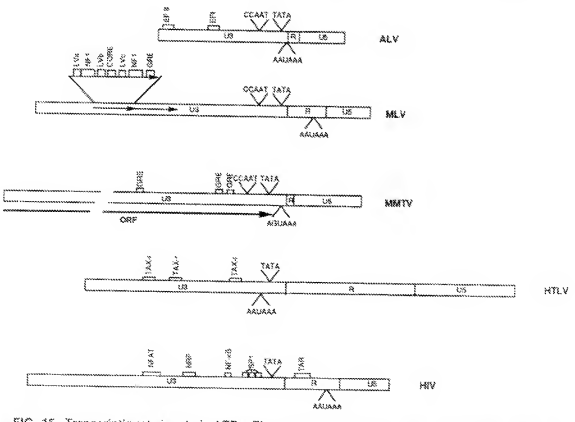

FIG. 15. Transcriptional signals in LTRs. The presence of specific sequence signals used for transcription and binding sites for known or suspected transcription factors is indicated for each LTR. (See refs. 369, 396, and 476 and Chapter 28.)

ECs relieves this block to MLV expression (246,247). Interestingly, expression of MLV proviruses introduced into undifferentiated cells is not induced by differentiation (305). This effect is presumably related to the permanent silencing of endogenous proviruses (see below), probably involving DNA methylation (330,404).

U3 sequences can also provide temporal (or conditional) regulation as well as cell-type specific regulation. Expression of MMTV in infected cells is strongly dependent on the presence of exogenously supplied glucocorticoid hormone (446), a dependence that is important to the transmission mode of this virus—from mother to offspring via milk. In general, MMTV expression in infected mice is confined to lactating mammary glands, consistent with the hormonal dependence (although other factors must be involved as well). This specificity can be traced, by mutational analysis, to the portion of U3 just upstream of the cap site (72,447), and it has presented a valuable model for analysis of the mechanism of hormonal regulation of gene expression in general. Other examples of this sort of regulation can be found in the human retroviruses and their relatives. HTLV-I expression is strongly enhanced by the action of the tax protein, acting through a region of U3 (108,124,212,311,441) [this regulatory ability can also be transferred into the MLV LTR (212)]; moreover, HIV expression appears to be positively regulated by the interaction with a factor like the NF-κB transcription factor, which is found in activated lymphocytes and has been postulated to effect activation of latent proviruses *in vivo* (235,295,441). These issues are covered in more detail in Chapter 28 and Reference 476.

Several lines of experimentation have been taken to dissect the specific sequences involved in the control of expression. Standard assays for enhancer and promoter function have been developed for this purpose. These assays usually involve measuring expression of the LTR linked to a reporter gene whose product is easily assayed [such as the bacterial chloramphenicol acetyl transferase (CAT) or β-galactosidase (306) after their introduction into appropriate cultured cells]. Using these techniques, it is readily possible to test the effects of rearranging, recombining, or mutating the LTR on its ability to direct expression. Results of such experiments are usually presumed to reflect the rate of transcriptional initiation (and, in general, seem to do so), but it should be kept in mind when reviewing them that they reflect a rather artificial situation in that the activity is being taken out of its context as an integrated provirus and is being introduced in much higher copy number than is the usual case. It has been repeatedly observed that expression of artificially introduced "proviral" DNA is much lower than that of the same DNA introduced as an integrated provirus by a standard infection. While the transfection technique has been quite successful, results should always be interpreted cautiously until they are verified in a more natural setting.

The genetic analyses of LTR function are nicely complemented by chemical and physical techniques that directly measure the binding of specific proteins (transcription factors) to viral DNA. These approaches, including gel electrophoresis retardation (or "band shift") assays and DNA footprinting methods, permit both (a) precise localization of important DNA sequences and (b) identification of the cellular activity responsible for binding them (192,309,369,396).

Taken together, the results of the genetic and biochemical analyses have permitted characterization of the LTR signals important for transcription. The following general conclusions have emerged:

1. With the exception of a few viral-coded *trans*-activators (such as *tax*), viral transcription control elements are acted upon by the same factors as are used by the cell for expression of its own genes.
2. Consistent with this, the viral signals have sequences similar to the normal cell counterparts.
3. These sequences are usually short (as few as 8 bases), and they are active in either orientation.
4. The factor recognition signals are often found repeated two or more times per LTR. Such repeats sometimes include only single elements; at times, extended regions of sequence containing multiple elements are duplicated. In some MLV strains, for example, a central region containing about 70 bases of U3 is often duplicated in tandem. This duplication has important effects on the ability of the virus to induce disease in infected animals (171).
5. The signals are often—perhaps always—present in very complex patterns and combinations. Presumably, combinatorial subtleties in the arrangement are crucial to the activity and specificity of the LTR. The potential for turning this organization to our advantage to create viruses with tailored expression and replication patterns is obvious, but the rules by which this game might be played are still largely unknown.
6. Small sequence changes affecting the signals themselves can have dramatic effects upon expression. Much of the U3 sequence, however, is not included in these signals and can be changed with only subtle or no change in expression (102,103,151,249,415).
7. The signals themselves can have quite different effects in different cell types. Several of the MLV sequences that activate transcription in fibroblasts and in other relatively highly differentiated

cell types can have a negative effect on expression in stem cells as well as in ECs.

The specific signals relevant to transcription initiation and found within U3 include the following:

TATA and CCAAT boxes. Sequences with a consensus suggested by the names are found in most LTRs (and in most other virus and cellular genes) in relatively fixed location upstream of the cap site. Their presence and relative location are indispensable for virus expression, and they are often seen to be strongly conserved among otherwise highly divergent LTRs. They cannot be deleted, inverted, or moved elsewhere without greatly reducing virus expression. Cellular factors interacting with these elements in cellular genes have been identified and, where studied, also found to interact with the virus counterparts (145,166,366).

Enhancers. Enhancers are defined as control sequences (or groups of control sequences), usually found upstream of the cap site, which increase the frequency of initiation of transcripts at nearby promoters but do not themselves specify or provide the sites for transcription. Within broad limits, they can be placed at different locations and orientations relative to the promoter elements they control. The most thoroughly dissected enhancer region is that found in the 70-base-pair repeat of the MLV LTR (Fig. 15). No less than seven binding sites for six different transcription factors can be identified (396); moreover, variation in the details of the precise sequence of this region, as well as the relationship of the sites to one another, is important in determining the level and cell specificity of viral expression and pathogenesis (35,50,51,103,137, 239,255,358,380,405,408). For example, viruses with only one copy of the enhancer region seem to be less capable of inducing thymic lymphoma in mice than those in which the region is duplicated (171). Indeed, rearrangements involving this region can be observed in animals during the course of infection and tumor induction by viruses containing only a single element (see below).

Corresponding regions of other LTRs have similar roles, but these regions differ considerably in the nature and arrangement of the factor-binding sequences. The HIV U3 region, for example, can be divided into four subregions, in order of increasing distance from the cap site: (i) the TATA and CCAT boxes; (ii) repeated binding sites for the ubiquitous cellular transcription factor SP1 (396); (iii) the site responsive to T-cell stimulation through the action of NF-κB (297,298); and (iv) a negative regulatory region that apparently responds to the negative effect mediated by the *nef* gene product (3,301) (see Chapter 28).

It is possible that in the context of a complete provirus, enhancers and other transcriptional regulatory elements may reside in locations other than upstream of the cap site. For instance, there is some evidence that the downstream LTR may contribute enhancer activity to the upstream one, since structures with two LTRs are more active in transcriptional assays than those with only one (307). Also, a sequence with enhancer-like activity has been reported to lie within the gag region of ALSV (7,49,202), but it has not been directly shown to be important in virus replication. These issues can be difficult to address experimentally, since modifications in the putative control regions also affect the structure of the transcript and thus can have posttranscriptional effects on processing, stability, or translation of the message.

Trans-*Activation*

In many retroviruses the regulation of expression is passive in that the only contribution of the viral genome to the expression process is the signals provided in the LTR and elsewhere which are themselves recognized by cellular factors. Viruses of at least two groups—lentiviruses and the HTLV-related viruses—are known to take a more active role. With these viruses, products of the virus genome are required to achieve high levels of expression. This phenomenon resembles that seen with a number of DNA viruses in which expression of virus genes is dependent on the presence of one or more viral gene products in the cell. In the case of both HIV and HTLV, it has been repeatedly demonstrated that LTR-driven expression is stimulated substantially by the presence of the relevant gene products (*tat* and *tax*) in the infected (or transfected) cells (156,441) (see Chapter 28 and Reference 476). Although the effect is very similar in the two virus groups, the mechanisms are quite different.

The HTLV tax protein seems to act as an accessory transcription factor. The target for its action is a 21-base sequence found several times within the U3 region (40,124,212,310,326,359,377) (see Reference 476). This sequence does not directly bind tax protein, implying that the effect of tax is indirect, presumably requiring its interaction with cellular transcription factors (365,377). In addition to its action on the expression of the homologous LTR, tax can also cause the transcriptional activation of a number of cellular genes—in particular, genes thought to be important in regulation of growth of the target T cell, including the T-cell growth factor, IL-2 and its receptor, the proto-oncogene *c-fos*, and others, such as the HIV LTR (via NF-κB) (11,32,365). This effect is thought to be important to the ability of these viruses to stimulate multiplication of these cells in culture and *in vivo*. Induc-

tion of growth factors and related genes has been postulated to play an important role in HTLV-mediated leukemia, although direct evidence is so far not available.

The HIV tat protein is unrelated to tax in terms of size, sequence, and mechanism (see Chapter 28). It is clearly not a usual sort of transcription factor. Its target (sometimes called TAR) is found in the R portion of the LTR and is thus present in the transcript as well as in DNA (157,290,361). This additional complication has made its mechanism of action quite difficult to study; several different possibilities have been proposed, including stabilization (or increased translational efficiency) of transcripts (360), bypassing of a transcriptional termination site (202,435), stimulation of transcription initiation (190,311), and combinations of mechanisms (111,157). Unlike enhancers, which don't seem to have significant secondary structure, the TAR region is characterized by a predicted stem-loop structure (at least in single-strand RNA), which may bind the tat protein directly (111). Effects of tat on expression of cellular genes have not been directly observed. Such effects are implied by the observation that mice carrying copies of an HIV LTR-tat construct in their germline suffer from a pathological condition resembling Kaposi's sarcoma, a common complication of AIDS (234,451).

There is a single report of a *trans*-activator-like activity encoded in the gag region of ALSV (42). Despite considerable effort, however, it has not been possible to verify the presence of such an activity using a variety of other approaches (48,96,307), and there is presently no good reason to believe that it is of importance to viral replication or transformation.

A special sort of *trans*-activation can occur between completely unrelated viruses (432). Expression of the *trans*-activating gene from a number of DNA viruses, including several herpesviruses (82,134,174,209,284, 285,316) and hepatitis B virus (375,381,436) and others (132), have been shown to activate expression from the HIV LTR, apparently by acting through several different mechanisms. The ubiquity of this phenomenon probably reflects (a) the relatively small number of different pathways available in the cell and (b) the consequent use of the same ones by unrelated viruses. It is easy to imagine how these mechanisms might have important pathogenic consequences in doubly infected individuals, but a direct demonstration of this has yet to be provided.

Negative Regulatory Factors

Those retroviruses which, like HIV, may have a biphasic lifestyle (latent alternating with lytic) might be expected to encode some sort of function which suppresses virus expression. Such a role may be served by the product of the *nef* gene, whose presence in cells leads to significant reduction of expression directed by the HIV LTR (3,8,301) (see Chapter 28). The effect of *nef* must be quite indirect, since its product is not found in the nucleus but, rather, is located near the inner surface of the plasma membrane and resembles a GTP-binding G protein in structure (117,160). It is tempting to imagine that the nef protein lies on a pathway for responsiveness of HIV expression to some external influence, but no such result has been reported.

3' Versus 5' LTR

The mode of synthesis of viral DNA enforces identity of sequence between the two LTRs, yet their functions in virus replication are quite distinct. While the use of transcription initiation signals to generate transcripts initiating in the 3' LTR is an important method of viral pathogenesis (see below and Chapter 13), it is a side effect of no obvious benefit to the virus. Indeed, close examination of RNA in cells infected with ALSV has revealed no trace of such transcripts (161). What suppresses the 3' LTR? It has been proposed, and supported with some experimental evidence, that the transit of the transcription machinery across the U3 region may inhibit its use as a site of initiation (77). This "promoter occlusion" mechanism has solid precedent in some other systems, but insufficient evidence is available to form a firm conclusion. Alternatively, there may be some signal adjacent to the 5' LTR which specifies its use for transcriptional initiation. The putative *gag* enhancer would be a reasonable candidate for such an activity, but no direct experimentation to test its role in replication or expression in a natural setting has been reported. Consistent with the idea of a 5' signal is the observation of 5' deletions, often not affecting the LTR, in proviruses that have activated expression of a proto-oncogene via transcription initiated in the 3' LTR (138,352).

Position Effects

It is frequently observed, especially by retrovirovectorologists, that the level of expression of a provirus can differ by a factor of 10 or more from one clone of infected cells to another. This effect is usually attributed to the variable influence of surrounding DNA on provirus expression. In support of a role of position in expression are the observations that integrated proviruses are expressed much more efficiently than unintegrated viral DNA and that proviruses acquired by infection yield higher levels of transcripts than do similar DNA molecules introduced by transfection techniques (181,321). The clearest evidence for

such effects, albeit in a special case, comes from analysis of MLV proviruses introduced into ECs in which the MLV LTR does not provide a usable promoter function (112,120,164,247,252,253,420,458). However, if cells that express viral genes are selected, a small proportion of surviving cells is found, and independent infections use nearby integration sites (15,421,458), suggesting the selection of rare cells in which integration had occurred in a region capable of directing expression of the adjacent LTR.

Processing of Viral Transcripts

Proviruses contain a single transcription unit; all genomic and mRNA species are derived from a primary transcript that begins at the cap site and extends through the provirus into flanking cell DNA. Responsibility for processing this molecule into finished genomes and mRNAs for the various proteins lies with cellular systems. Again, with many retroviruses the virus provides only signals; HIV and HTLV and their relatives take a more active role in part of the process.

Polyadenylation

With all retroviruses the genomes and mRNAs have identical ends: (a) the 5′ end as determined by the site of initiation of transcription and (b) the 3′ end at the R–U5 border set by the site of poly(A) addition. The polyadenylation reaction seems in all important respects identical to that used by the cell to process transcripts of its own genes (24). Although primary retroviral transcripts have not been studied, they most likely extend well beyond the eventual 3′ ends. Specific signals in the transcript, including the hexanucleotide AAUAAA, specify coupled cleavage and poly(A) addition reactions, commonly following a CA sequence 12–20 bases downstream. An interesting exception to this virtually universal pattern is found in the greatly larger separation of the presumed poly(A) signal and the poly(A) in HTLV and BLV. It has been postulated (but not yet proven) that most of the sequence between the putative poly(A) signal and its site of action can form a highly ordered stem-loop structure that effectively moves the poly(A) site into the correct relative position (372).

An interesting problem with understanding retrovirus expression stems from the presence of identical poly(A) signals in both LTRs. As with transcriptional initiation, why is only the correct site used and not the other? With ALSVs and some other retroviruses, the answer is obvious: The AAUAAA poly(A) signal is in U3 and is therefore present only once in the transcripts. With most others, primary transcripts have two copies of this sequence—one in each R region. It has been suggested that some sequence upstream of R in the transcript (and therefore found only at its 3′ end) may be required for efficient polyadenylation; moreover, some experimental support for this idea, in the form of U3 deletions that apparently affect the addition of poly(A) to transcripts, has been obtained (93).

In the natural case, polyadenylation at the correct site is not always accomplished with perfection. In cells infected with ALSV, some 15% of transcripts contain complete LTR sequences followed by up to 2 or more kilobases of additional nonviral sequence at their 3′ end (162). Since they are polyadenylated, these RNA molecules are most likely to have resulted from poly(A) addition to a long primary transcript at sites derived from host-cell sequence flanking the provirus. Although they are most probably irrelevant to normal retrovirus replication, such molecules are biologically active. They can be incorporated into virions and serve as functional genomes (419). They may have played an important role in the acquisition of oncogenes in some retrovirus groups (163) (see below).

Splicing

The single primary, polyadenylated transcript serves to provide genomes and all necessary mRNA species (Fig. 16). All retroviruses use full-length transcripts for genomes and mRNA for the *gag* and *pol* genes, and all yield, in addition, at least one spliced mRNA, which encodes env. In HTLV and lentiviruses, more complex splicing patterns give rise to as many as six subgenomic species encoding not only env but also the various regulatory proteins (4,107,291,373, 392,457). In all retroviruses, the spliced RNAs share 5′ and 3′ ends with the genome, and all subgenomic RNAs in a given virus have a splice starting at the common 5′ donor. In most virus groups, this site is located in the leader region upstream of gag, and genes encoded by subgenomic species supply their own initiation codon. There are some exceptions to this. In HTLV the donor is located in R about 130 bases from the 5′ end; in ALSV it is found within gag such that the complete leader and the first six codons of gag are appended to the 5′ end of smaller mRNAs (447).

The necessity for some sort of control of the splicing process is obvious: An appropriate balance between amounts of the various proteins and genomes has to be maintained for optimal replication rates. How this balancing is accomplished is, at present, poorly understood. With most—perhaps all—retroviruses, regulation appears to be passive, and splicing is not directly affected by viral gene products themselves. It follows then that signals in the transcripts must specify the partial use of their splicing signals. Sequences of the gene expressed by a given mRNA itself are probably

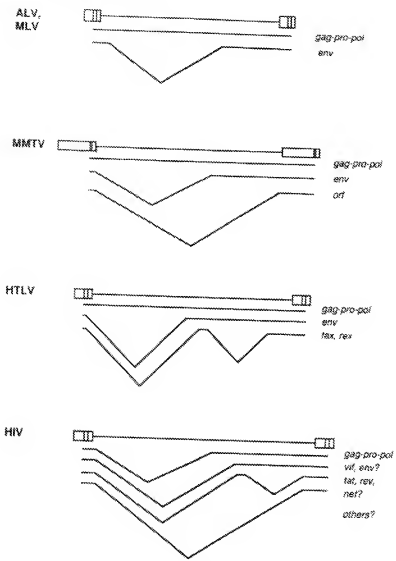

FIG. 16. Splicing patterns of some retrovirus mRNAs

not important; experience with several types of retroviral vectors has shown that many eukaryotic and prokaryotic genes can be placed downstream of splice signals without greatly affecting splicing patterns (62). The sequence of the identifiable splicing signals themselves also does not seem to play a major role. To the extent that consensus sequences can be identified, they closely resemble their normal cell counterparts; moreover, the acceptor sequences used by some oncogenes, such as *src*, have been incorporated into the genome essentially unaltered from the ancestral proto-oncogenes (437) but now display the splicing pattern characteristic of the virus.

These considerations suggest that sequences affecting the extent of splicing should be sought within the portion of the RNA that serves as intron (i.e., within the gag and pol coding regions). Examination of these regions has indeed revealed effects of the expected sort: Many different mutations introduced into them have strong effects on the relative levels of spliced and unspliced RNAs (6,205,280,407,409,410). Unfortunately, the effects observed are very complex, and no one specific region or structure seems to be implicated. For example, insertion in the ALSV genome of a short sequence 12 bases upstream of the splice acceptor for *env* causes the acceptor to be used much more efficiently and leads to loss of infectivity of the resulting virus (205). Selection of replication-competent virus is accompanied by the appearance of a variety of single-base mutations in the general vicinity of the duplicated sequence, but not necessarily directly affecting either it or the consensus acceptor sequence. Other mutations affecting splicing have also been found in the *gag* region of the same virus (6). So far, it has been impossible to provide a mechanistic interpretation for these effects, beyond general statements that the overall structure of the intron is in some way important to achieve the correct balance. The issue is an important

one for the design of viral vectors, since it is often desirable to use viruses that express two genes in the manner of the normal viral genes; there is presently no way to predict whether a given sequence in the first position will be compatible with expression of the second.

In principle, regulation of splicing could be accomplished either by direct effects on the splicing machinery itself or by affecting transport of unspliced species (i.e., by causing them to be moved out of the nucleus before splicing can occur). While the former mechanism would seem more likely with most retroviruses, there is evidence for regulation at the level of transport or stability with regard to the HTLV group and the lentiviruses. Both these viruses encode gene products (*tax* and *rev*, respectively) which, although quite different in size, structure, and mode of expression, affect relative mRNA levels in a similar fashion (106,107,149, 182,263,363,389), possibly through the same pathway (348). These proteins act to regulate the relative mRNA levels in a particular way. In the presence of rex or rev, the full complement of mRNAs is generated. In their absence, only multiply spliced mRNAs accumulate in the cytoplasm; unspliced and singly spliced *env* mRNAs are found at reduced levels and only in the nucleus. The inferred mechanism for this effect is that the *env* gene contains a sequence that does not permit transport from the nucleus unless the corresponding protein is present; moreover, interaction—quite possibly indirect—of a sequence in *env* with the regulatory protein relieves the block (149,362,372). The HIV sequence—known to one group as RRE (for rev-responsive element)—is a region of RNA predicted to have a complex secondary structure (262). Since it confers responsiveness to *rev* activation on RNAs (derived from synthetic constructs) that cannot be spliced, its effect seems to be due to its ability to activate transport of RNA from the nucleus. This implies the existence of some mechanism to prevent transport to the cytoplasm of RNAs that are not fully spliced, and suggest that a similar mechanism—the provision of a special signal for transport of incompletely spliced RNAs—may also be important in other retroviruses, even in the absence of a virus-coded factor to regulate the process. Proteins regulating relative amounts of spliced mRNAs could feed back into the overall replication cycle in important ways.

Since the *trans*-activating proteins—tax and tat—are expressed from multiply spliced RNAs, they would accumulate early after infection or early after induction of expression. Later, as the regulatory proteins increased in amount, relatively more virion proteins would be synthesized. Thus, a sort of temporal regulation of expression could be achieved, crudely analogous to that seen with many DNA viruses.

Translation

After transport to the cytoplasm, the spliced RNAs serve as mRNA for *env* as well as for regulatory genes, or some oncogenes when present. The full-length RNA has two different fates. Some molecules become new genomes; others serve as message for the several proteins encoded in the *gag*, *pro*, and *pol* genes (448). How molecules are selected to enter either the mRNA or the genome pool is not known. Perhaps initial binding of the gag precursor is sufficient to repress translation and to commit an RNA to becoming a genome. When the concentration of a gag protein is low, RNA would instead become associated with polysomes and would subsequently be unavailable for packaging. This question is intimately associated with the issue of packaging of genomes and deserves further study.

The gag, pro, pol Protein Product

Compared to cellular mRNAs, retroviruses have unusually long leader regions preceding the initiation site for translation of gag (see Table 2). Unlike picornaviruses, which seem to use a special mechanism for initiating protein synthesis at translational initiation sites located at some distance from the 5′ end of the genome (see Chapter 16), the translation of retrovirus mRNA apparently follows the standard "scanning" model in which ribosomal subunits bind initially to the capping group and move along the RNA until the gag AUG initiation codon is encountered (333). At least in the case of ALSV, this means traversing several small open-reading frames, each of which has an initiation codon in a less-than-optimal context but at least one of which is translated into a small peptide (148). It has been speculated that translation of these sequences has the function of aligning the ribosomes appropriately for the major task of translating gag (148), but direct evidence favoring this interpretation over random chance is not yet available. The ALSV leader seems to be unusually specific for the class of host cell. In mammalian cells, it is utilized very inefficiently (206). This type of specificity is not observed in other retrovirus groups; certain murine leukemia viruses can replicate efficiently in avian cells (459).

An unusual feature of some mammalian C-type viruses is the presence of two *gag* gene products. In both murine and feline leukemia viruses, translation of a significant fraction of RNA molecules initiates at a codon well upstream of, and in the same frame as, the standard *gag* initiation codon (85,86,437). The resulting

protein thus contains all of the gag sequence with additional amino acids at its 5′ end. The additional amino acids contain a signal peptide that directs membrane translocation and then is removed. The resulting uncleaved gag precursor is then processed like an extracellular protein. It is glycosylated and exported from the cell and ends up in the extracellular matrix (334). Exactly where translation of this protein is initiated has not been established. It is likely that, consistent with the "guidelines" for ribosome scanning, the initiation codon for the glycosylated gag protein is a relatively unfavorable match to the standard consensus and, thus, that it is often skipped by ribosomes so that the standard AUG codon can be used. Indeed, in MLV, there is no AUG in the appropriate position, and it has been suggested that a GUG codon, usually encoding valine, is occasionally used for this purpose (C. Van Beveren, *personal communication*). Initiation at codons other than AUG is not known in eukaryotic systems, but it has been seen in a similar context translation of some bacteriophage genomes. The function of the glycosylated gag protein remains mysterious. Although its conservation implies a functional role, introduction into the MLV genome of mutations expected to prevent its synthesis have no discernible effect on the replication or pathogenesis of the virus (85,86).

In all retroviruses the *gag*, *pro*, and *pol* genes can be viewed as forming a single translational unit, which is expressed to yield a set of nested primary translation products of the sort

gag
gag—pro
gag—pro—pol

such that the amount of pol protein produced is only one-tenth to one-twentieth that of gag protein. (Note that most viruses encode only one or the other of the first two products.) This strategy serves both to ensure a proper ratio of the various proteins to one another and to provide the pro and pol proteins in association with gag so that they can be incorporated into virions (see Fig. 6). There are two different mechanisms that retroviruses have evolved to accomplish this proportional synthesis, both of which take advantage of quirks in the eukaryotic translation machinery and involve provision of a translational termination signal at the gag-pro and/or pro-pol boundary that is occasionally bypassed by the translating ribosomes to incorporate the adjacent downstream protein into the same molecule.

The first such case to be recognized was that of MLV (Fig. 6). In this instance, the *gag* and the *pro-pol* reading frames are in the same reading frame but are separated by a single UAG (amber) translational terminator (437). Examination of the amino-terminal amino acid sequence of the mature PR protein revealed the presence of a glutamine residue, usually encoded by a CUG (or, less likely, CUA) codon (474,475). Fortunately for this analysis, the amino-terminal end of PR, generated by cleavage from NC, is 4 bases upstream of this site. Thus, approximately one time in 20, the translational machinery must misinsert a glutamine at the position of the terminator, allowing continued translation through pro and pol. This sort of nonsense suppression is a well-known mechanism in bacteria, but it is not known to occur in normal eukaryotic gene expression. It has been observed in the expression of the nonstructural proteins of alphavirus genomes, where it apparently plays a similar role (414).

What causes nonsense suppression is not entirely clear. A rare glutamine tRNA with the ability to mediate the suppression has been observed, and it has been suggested that its expression may be induced by MLV infection (222), although infection is not required for the phenomenon (110). In general, UAG terminators are not suppressed in eukaryotic cells, suggesting that the context of the sequence might be important in specifying the event.

In all other retrovirus groups, the *gag* (and/or *gag-pro*) reading frame ends in a translational terminator and is also in a reading frame different from that of *pol* (Fig. 6). Thus, in the majority of cases, simple nonsense suppression cannot suffice for readthrough into the next coding region: a shift of reading frame must occur as well. Again, there is precedent for such a mechanism in certain prokaryotic systems, but none had been known in eukaryotes; moreover, the existence of a specific spliced mRNA was once considered the more probable mechanism (61). Clear demonstration that the mechanism of readthrough involved "frameshift suppression" was obtained initially for the junction between the *gag-pro* and *pol* frames of ALSV when RNA which was synthesized *in vitro* by purified prokaryotic enzymes was used as an mRNA *in vitro*. This RNA, which could not possibly have been spliced, could still serve as mRNA for the appropriate ratios of the two polyproteins when presented to eukaryotic ribosomes (187). Using this system, it was possible to define precisely the sequences necessary for the phenomenon and to determine the amino acid sequence at the site of frameshifting (185). In the ALSV case, the following sequences were found:

```
PR—...Leu Thr Asn Leu   ***
        UUGACAAAUUUAU AGGGAGGGCC...
RT—...Leu Thr Asn Leu
                  Ile Gly Arg Ala...
```

Here there is apparently an occasional slippage of the ribosome to the −1 frame so that the sequence around the terminator is read AUA-G . . . instead of the expected . . . A-UAG. This slippage requires the

seven nucleotides shown in boldface as well as the presence of a region with a predicted stem-loop structure just beyond the frameshift site. These requirements seem to be general. In all viruses studied, with the possible exception of HIV (467), a specific (but different) seven-nucleotide A-U-rich sequence at the frameshift site, along with a stem-loop structure following it, seems to be the essential feature for shifting of translation to the −1 frame (73). Apparently the specific sequence permits (but does not force) the ribosome containing the charged tRNA in the A site to slip back one base when it encounters the obstacle formed by the secondary structure. The efficiency of this event is adjustable over a wide range so that similar ratios of gag and pol products always result, regardless of the configuration. In ALSV, where only one such event is required, the probability of frameshifting is about 5%. In MMTV [as well as in HTLV-2, where *gag*, *pro*, and *pol* are in three different reading frames requiring two frameshifts (Fig. 6)], the first takes place about 25% of the time while the second takes place about 10% of the time, giving an overall ratio of gag-pro-pol precursor to gag precursor of about 2.5% (169,187,262). In the case of HIV, it has been reported that the "slippage" signal alone is sufficient to induce frameshifting from gag to pro in the absence of the stem-loop structure (467).

Again, this mechanism has not yet been seen in eukaryotic gene expression, but it is used in at least one other virus group—the coronaviruses (41; see Chapter 18). It is also apparently used in at least one retrotransposon, the yeast Ty1 element (468). In this case, however, the short recognition sequence alone (in the absence of any special secondary structure) is apparently sufficient to effect the shift—in this case, in the +1 direction. Nevertheless, the principle of the reaction seems to be the same: A specific sequence that permits the shift is combined with a feature that delays translation momentarily to force it to occur (P. Farabaugh, *personal communication*).

The env Protein Product

Unlike the translation of *gag*, *pro*, and *pol* which takes place on free polyribosomes using full-length RNA, the env protein product, like other cell-surface proteins, is synthesized on polyribosomes associated with the rough endoplasmic reticulum (ER) using a spliced subgenomic mRNA (85,86,447). To direct this process, all primary env protein products contain an unremarkable signal peptide at their amino termini. As with cellular proteins, this peptide directs the system to initiate transmembrane synthesis and is itself cleaved off after insertion into the rough ER by the normal cell machinery for this process. Following translation, all retroviral env precursor proteins remain anchored to the membrane by the hydrophobic membrane-spanning domain near the carboxy terminus; deletion of this sequence causes the protein to be exported and released from the cell in soluble form (329,466). The first reaction in the addition of N-linked carbohydrate side chains occurs during or shortly after translation. Further modification of the carbohydrate (trimming of the high-mannose core, addition of other carbohydrates to the branches, addition of terminal sialic acid residues) takes place in the Golgi apparatus, as does cleavage of the precursor into the SU and TM peptides, which remain associated with one another (85,86). The cleavage uses a cellular enzyme system, presumably the same one used to process the influenza HA protein as well as other viral surface proteins (see Chapter 4). It always occurs at a characteristic amino acid sequence—following a string of at least three of four basic amino acid residues, usually of the form Arg/Lys-x-Lys-Arg. Failure of this cleavage to occur does not necessarily inhibit further processing or assembly but does yield noninfectious virions (272,332). In MLV and related viruses, a final modification of the *env* gene product is the removal of a small C-terminal fragment (sometimes called the "R" peptide) from the end of TM. This cleavage occurs during budding and is probably catalyzed by PR (86).

Virion Assembly

Assembly of virions is a poorly understood aspect of retroviral replication. In the electron microscope, two patterns can be discerned, differing in the site of assembly. With most virus groups, assembly of capsids and budding are simultaneous. The assembling particles are first visible as crescent-shaped patches on the inner face of the cell membrane, which then appear to extend until the ends meet to form a hollow sphere as the membrane wraps around to form the envelope. With B- and D-type viruses, capsid assembly takes place in the cytoplasm, leading to formation of an A-type particle, which only later associates with the membrane and buds out. In both cases, a structural rearrangement of the capsid to a more condensed form is visible during or shortly after release of the virion.

A schematic diagram of how this process is currently thought to occur, based on a model first proposed over 10 years ago (33), is shown in Fig. 17. This scheme proposes that assembly involves interaction of the three principal domains of the gag precursor protein: at its C-terminal (NC) end with the genome; in the middle (CA) with the CA region of other precursor molecules; and at the N-terminal (MA) end with the cell membrane. In the case of B- and D-type viruses, the membrane interaction would take place only after

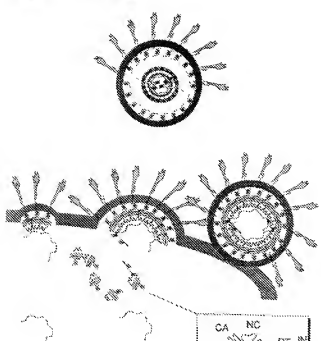

FIG. 17. A model for retrovirus assembly. The scheme would apply to most retroviruses (except B and D types). Note that cleavage of proteins and assembly of the genome do not take place until around the point of release of virus.

the others; in all other viruses, the three interactions would be more or less simultaneous. Thus, the shape of the virion is formed by the CA protein interactions, the genome is incorporated through interaction between it and the NC domain, and the envelope is acquired as a consequence of the interaction between the MA domain and the cell membrane. The same interactions on the fraction of molecules extended through pro and/or pol serve to bring these domains into the virion in the appropriate location and amount. Following assembly, proteolytic processing mediated by the PR protein separates the domains into individual proteins, thus preparing the system for reverse transcription when it encounters the proper environment. The separately processed and exported env protein complex is incorporated into the virion envelope via some sort of interaction with one of the gag proteins, presumably MA.

This scheme has a few interesting features. First, the capsid proteins function in different forms at different times in the life cycle—in a polyprotein precursor during virion synthesis, and as cleaved proteins during early events (penetration, reverse transcription, and integration). Second, the process is unidirectional. It would probably be impossible to assemble a capsid from the cleaved components, since it is not assembled that way in nature. Also, the unprocessed form of the capsid is most likely inactive for reverse transcription [although the MLV gag-pro-pol precursor has reverse transcriptase activity (109)], preventing intracellular buildup of viral DNA, a potentially lethal event for the infected cell (see below). Third, although all proteins and the genome are required for full infectivity, most are dispensable for assembly of recognizable virions. Indeed, virions are produced in the absence of (a) genomes, (b) the pol, env, and pro products, and (c) the NC domain of gag. The only two domains required for the process are MA and CA.

Role of Specific Proteins

Some genetic analysis, as well as a limited amount of physical analysis, has been applied to determine the role of the various domains of the precursor proteins in the formation of retrovirus virions. In general, the analyses support the model presented above and in Fig. 17 and illuminate some of the details of the process.

The MA Protein

In most retrovirus groups, the amino-terminal domain of gag is modified by the addition of a hydrophobic myristate chain. The expectation that this modification might be important for membrane association is well supported by experiment. Mutations in MLV which eliminate the ability of MA to be so modified also eliminate membrane association, capsid formation, and budding (71,197,344). Interestingly, such mutations also eliminate cleavage of the gag polyprotein (368), consistent with the idea that cleavage requires assembly. Amino acid alterations away from the short amino-terminal sequence that signals the myristylation site do not inhibit budding but yield noninfectious virions (75). The specificity of the myristylation requirement is better demonstrated in D-type viruses, where a myristylation-negative mutant exhibits no defect in capsid assembly but exhibits, instead, only membrane association and budding (346). The ALSV MA protein is an unusual exception to the rule of gag myristylation. Its amino-terminal acetyl group can hardly be sufficiently hydrophobic to support membrane association, nor does it possess a significant stretch of hydrophobic amino acids that might fill this role. To date, there is no good hypothesis to explain the membrane association of this protein.

At one time it was thought that the MA protein was also an RNA-binding protein; indeed, that specificity for recognition of the genome RNA resided in it (86). More recent evidence suggests that the original observations might have failed to distinguish completely

among the gag proteins (272), and current thinking assigns this role to NC.

The final role attributed to MA is interaction with the env protein complex. It seems a truism that specific association between MA and some portion of the TM protein must direct incorporation of the env complex into virions, but there is no direct evidence demonstrating the association of the regions involved.

The CA Protein

There is no reason to doubt that CA forms the core shell which is prominent in electron micrographs and which is necessary for formation of capsids (85,86). Mutations introduced into the MLV CA protein gene result in a complete failure of assembly (177). Beyond this, little is known about the structure or mechanism of action of this protein in virion synthesis.

The NC Protein

The NC protein, as befits its RNA-binding properties (85,86,121,122,272), has been shown to be required for inclusion of RNA in genomes. In both ALSV and MLV, mutations in the region encoding the NC domain, particularly in the Cys-His boxes supposed to mark binding sites, lead to synthesis of virions lacking genome RNA and, in some cases, possessing no RNA at all (140,273–275). Although most binding assays characterize NC proteins as nonspecific RNA-binding proteins of high affinity, this protein is most likely to confer the necessary specificity for packaging genome RNA as well. This issue has not been well addressed with the reagents tested to date. One could imagine that a level of specificity might exist in the uncleaved precursor but might then be lost from the cleaved form; alternatively, otherwise modified variants—for example, variably phosphorylated forms (122)—may also be important in permitting both specific and nonspecific interactions.

Genetic and biochemical analysis of the ALSV NC protein reveal a second possible role for this protein in assembly: mediation of RNA–RNA interactions. Certain mutants defective in the Cys-His boxes of the protein package genome RNA but do not support its dimerization (274). Annealing studies show that addition of NC to mixtures of short complementary RNAs (like the primer tRNA and its binding site) hastens the rate of duplex formation to a considerable extent (338). This activity could be quite important in catalyzing the formation of short duplex structures (such as that between the primer and its binding site or between the two subunits) that might not easily form by themselves.

The PR Protein

The protease domain is not required for assembly or budding: Mutants lacking activity synthesize virions that (a) are noninfectious, (b) contain only uncleaved precursor, and (c) do not exhibit morphological maturation (204,215,453). These results imply that the PR protein is not active until a late stage of viral maturation and that it must not be freely present in the cytoplasm of the infected cell. Premature cleavage of gag would obviously be deleterious to virion synthesis. Consistent with the requirement of assembly for protease activity is the lack of self-processing of (a) the ALSV gag-pro precursor *in vitro* and (b) the MLV precursor *in vivo* when assembly is blocked by mutations in the myristylation site (368). Thus, the PR domain in the precursor form must have little or no activity until the PR domain itself is cleaved out. At one time it was supposed that another cellular protease might be necessary to initiate the process; it now seems more likely that some special feature of virion assembly activates the enzyme, since precursors containing the ALSV or HIV protease have been shown to undergo spontaneous cleavage when expressed in bacteria or yeast (104,217,220,271,277,287). Conceivably, the activation is a simple matter of attaining a sufficient critical mass. If the enzyme is very slightly active in precursor form (and then only in *trans*), no cleavage could occur until a very high local concentration of protein formed in the budding virion. Once the initial cleavage(s) occurred, active enzyme molecules would be released which would cleave out more PR proteins, leading to a rapid chain reaction and virtually instantaneous processing of the entire capsid. A variant on this idea is provided by the recently deduced structure of the HIV and ALSV PR proteins (279,297,384,457). Both proteins were found to be dimers joined in such a way that the dimer linkage region contains the active site. If, as this suggests, the proteases are only active as dimers, then such dimers might not be able to form until the precursors have aligned themselves properly as in the late stages of budding.

The pol Proteins

Mutants lacking pol proteins assemble, bud, and process their capsids normally, but are, of course, noninfectious. Because at least some RT proteins exhibit specific binding of the "correct" tRNA primer and because mutants of ALSV defective in *pol* do not incorporate the tRNA into virions (446), it has been supposed that binding of RT to tRNA is an important mechanism for ensuring inclusion of the primer in virions.

The proportion of pol-containing precursors is also

apparently critical. Introduction into the MLV genome of a point mutation altering the gag terminator to a glutamine codon, so that all molecules synthesized resemble the 5% of suppression products seen in the normal case, causes a failure of assembly to occur, perhaps because the core structure cannot accommodate a large number of RT molecules (109).

The env Proteins

As with pol, the env proteins are not necessary for virion assembly or budding. In their absence, normal-appearing virions (except for the absence of surface projections) are made and released. Such virions are also infectious if they are introduced into cells by artificial means, implying that env proteins are required only for initial events (adsorption and entry) in infection (86).

As noted above, the signals that specify incorporation of env proteins into virions are a mystery at present. *A priori*, it seems reasonable to suppose that the usually short C-terminal cytoplasmic domain of the TM protein specifically interacts with the capsid. Surprisingly, however, deletion of this entire region from the ALSV TM protein has no detectable effect on the synthesis, processing, or incorporation into virions of the env proteins (329). Conceivably, the key interactions are with amino acid sequences within the transmembrane domain or on the extracellular side of TM, but it is difficult to see how the ALSV MA domain could interact in this way. That the process may not be tightly specific is implied by the ability of MLV to incorporate into functional virions, as well as to utilize, env proteins from unrelated retroviruses such as HTLV-1, albeit at low efficiency (469). Clearly, this issue deserves more study.

The RNA

In all retroviruses studied to date, the principal determinant of packaging on the genome is in the packaging signal in the leader region between U5 and gag. Deletions in this region do not affect expression, translation, or assembly but do result in the production of virions devoid of genome RNA (19,208,244,245,265, 293,303), often containing other RNA molecules either from related endogenous proviruses or from adventitious cellular sequences (101,127). The sequence so identified as essential for packaging has been termed Ψ (267) or E (456). In no case has the minimum essential sequence been precisely defined, and in only one case has a sequence sufficient for packaging been described. The region of the MLV genome extending from just downstream of the splice donor to immediately preceding the gag initiation codon can permit packaging of subgenomic MLV RNA if placed at other locations, or even of unrelated RNAs containing no other viral sequence (264). For full efficiency, however, a longer sequence, probably extending well into *gag*, is required (1).

The results with the mammalian C-type viruses explain the selectivity of packaging for genome RNA, since the packaging signal is absent from the spliced mRNAs. More difficult to explain is the situation in ALSV. Since the ALSV splice donor is within *gag*, spliced mRNAs contain the entirety of the signal identified so far. Subgenomic ALSV RNAs are also discriminated against during packaging, but perhaps not as effectively, since there are a number of reported cases of proviruses derived from reverse transcription of mRNA molecules that must have been incorporated into virions (29,416). Since the minimum sequence sufficient for packaging has not been identified, it remains possible that unidentified essential sequences lie within *gag*. Such a sequence, if it exists, must lie very near the 3′ end of *gag*, since vector constructs that include only 50 codons of *gag* are capable of being packaged (306).

In two groups of viruses, additional regions of the genome have been identified as necessary for its incorporation into virions. In MLV, deletion of a sequence near the 5′ end of U5 reduces efficiency of packaging substantially (293). This result is paradoxical, since U5 is not required for packaging foreign RNAs (1). Perhaps the U5 signal is specific for MLV RNA—for example, to enable correct folding. In ALSV, a 150-base sequence near the 3′ end of the genome has similarly been assigned a role in packaging (395); whether this is a private signal for virus genomes or a general requirement remains to be seen.

Except for size, there seems to be little in the way of additional requirements for packaging of RNA. To date, a very large number of different sequences have been passaged in retrovirus vectors. No failure of an RNA sequence to be packaged has been reported, implying that the constraints for packaging of any sequence are few. The precise 3′ end of the RNA is also unimportant; genome RNA molecules extended into cell-derived sequence at their 3′ ends are efficiently incorporated into infectious virions (417).

The most severe limitation on packaged RNA is that imposed by size. No lower size limit has yet been reported; the smallest ALSV known to be transferred is about 2.2 kb (23); the smallest MLV about 3 kb (62). The largest RNA reported to be transferred in a retrovirus virion is just over 10 kb (306); in several other studies, RNAs larger than about 10 kb were found not to be packaged into infectious virions (131). Although these experiments were not done systematically, they suggest a strong constraint. The level of this restriction—whether on physical assembly into virions, on

virion structure or stability, or on ability of the genome to be reverse-transcribed—remains to be determined. The observation that 3′ extended genomes as large as 11 kb can be seen in virions raises the possibility that the restriction might be at some level other than virion assembly (162).

SPECIAL FEATURES OF RETROVIRUS BIOLOGY

Host-Cell Effects

The retrovirus replication cycle does not require that the infected cell be significantly harmed or otherwise affected by the virus-specified events; moreover, the majority of retroviruses have little or no effect on the infected cell, other than to render it permanently capable of producing virus at a low level. However, the unique association retroviruses share with the host cell leads to numerous special features of retrovirus biology, and it results in a remarkably wide range of effects on the host organism. These effects range from malignant or degenerative disease through benign viremia to short- or long-term genetic effects such as insertional activation or inactivation of cellular genes. Most of the effects discussed here are considered in more detail in specific chapters and will be presented only briefly.

Transformation

Many retrovirus groups first came to light as agents which were isolated from naturally occurring tumors and which were capable of causing such tumors in appropriate hosts (26,27). It soon became apparent that these isolates could be divided into two distinct groups: one that causes tumors rapidly (1 week to a few weeks) and causes readily visible transformation of appropriate target cells in culture, and one that induces malignant disease only slowly (with a typical latency of 6 months to a year) and has no visible effect on cells in culture. Generically, the first group of viruses is often referred to as "transforming" retroviruses; the second group is often called "leukemia" or "leukosis" viruses (ignoring the fact that many of them cause disease other than leukemia), or even "nontransforming" viruses. The molecular distinction between these groups is now very clear. The transforming viruses act on cells through the expression of a specific oncogene unrelated to viral sequence; the leukemia viruses lack such a sequence and transform cells by other means.

Viral Oncogenes

Alone among animal viruses, retroviruses have the ability to incorporate fragments of certain cellular genes (known as c-*onc* genes or proto-oncogenes) into their genome and to alter their structure and expression in ways that enable them to directly transform a normal cell into a malignant one. Over two dozen different viral oncogenes (v-*onc* genes) have been found in retroviruses (some of them many times over), and study of these has been an important and rewarding aspect of fundamental cancer research over the last two decades. Detailed analysis of their structure and function as well as their relationship to their normal counterparts has provided insights into molecular mechanisms of oncogenesis, growth control, and signal transduction to a depth impossible to match in any other system. Discussion of these aspects is found elsewhere (e.g., Chapter 13); here I describe a few general aspects of the association of these genes with the viruses that carry them and the cells they transform.

All retroviral oncogenes are very recently derived from the genome of the host—most likely within the animal from which the transforming virus was isolated. There is no evidence that transforming viruses are efficiently passed from one individual to another, and there is good reason to believe that they are not. Retroviruses, in general, are transmitted horizontally very inefficiently, and transforming viruses kill the infected host very rapidly. The most productive sources of new viral oncogenes have been populations of animals in which large numbers are routinely screened for tumors and within which leukemia virus infection is common. In practice, these have been (a) chickens raised for slaughter and (b) pet cats presenting to veterinarians with tumors. Although the events that give rise to oncogene-containing viruses are probably very rare, the selection is powerful, since small numbers (perhaps one) of events per animal are amplified into a visible effect. Nevertheless, oncogene-containing viruses cannot be considered to be natural infectious agents. Rather, they should be thought of as rare aberrations that, if not given a good home in the laboratory, would die with the animal in which they arose.

Viral oncogenes differ from their cellular antecedents in a number of important ways. First, they often contain only a portion of their corresponding proto-oncogenes, limited to a subset of the region transcribed into mRNA. Second, they are derived from processed versions of the transcripts, with all internal introns removed by splicing. In some cases (such as *src* in RSV), a small fragment of intron along with the 5′ splice acceptor site remains and provides an acceptor for processing of the subgenomic v-src mRNA. Third, they have been separated from the cellular controls on their expression, including both the normal cellular promoter as well as other controlling sequences such as those which confer instability on the corresponding mRNA. Their expression has come completely under the control of the viral LTR and, thus, can differ from

that of the proto-oncogene in level, in lack of regulation (by growth factors, for example), and in cell type specificity. With some viral oncogenes, such as *myc* and *mos*, differences in regulation of expression seem to be sufficient to convert a proto-oncogene into an oncogene. Fourth, many oncogenes have suffered deletions or other rearrangements of sequence affecting the structure of the protein product. For example, v-*src* differs from c-*src* most prominently by the replacement of a short amino acid sequence at the very C-terminus of the protein. This difference includes loss of a tyrosine residue, the phosphorylation of which apparently regulates the tyrosine-specific protein kinase activity of the c-src protein. Similarly, v-erbB differs from its cellular relative (now known to be the receptor for epidermal growth factor) by the deletion of a large amount of sequence, amounting to most of the extracellular region including the ligand-binding site. Changes such as these are presumed to have the effect of allowing a signal-transducing function of the proteins to occur independently of the normal signals. Finally, viral oncogenes are often joined to viral genes in ways important for modifying their function. For example, v-*abl* is expressed as a fusion protein with a portion of the MLV gag protein as its amino terminus. This fusion provides to the gag-abl protein the gag myristylation site, and it directs the protein to a membrane site essential for its activity. The *fms* oncogene (derived from the receptor for the growth factor CSF-1) is also expressed as a fusion with the amino-terminal portion of the feline leukemia virus *gag* gene. In this case, the important feature is the signal peptide provided by the glycosylated form of the gag protein, which allows proper placement of the protein into the cell membrane.

With the exception of some strains of RSV, virtually all oncogene-containing retroviruses are defective for replication because the oncogene has replaced all or part of the essential protein-coding region of the genome but retains all necessary *cis*-acting sequences. Nevertheless, such viruses are capable of replicating as mixtures with the corresponding nondefective, nontransforming virus. In this context, such a virus is known as a "helper" virus. In a doubly infected cell containing proviruses of both the transforming and the helper virus, the proteins synthesized by the helper can provide all the functions necessary to encapsidate the defective virus genome and allow it to generate an integrated provirus in the next round of infection. This only works because retroviruses capable of acquiring oncogenes do not contain *cis*-acting signals for expression within the coding region, nor do they encode *trans*-activating or other regulatory proteins necessary for their expression. Cell lines containing only the transforming provirus can be obtained by (a) cloning infected cells shortly after infection with a mixed population of virus and (b) testing for the absence of virus production. Such cell lines are called "nonproducer cells" and have been quite useful in studying oncogene function.

Insertional Activation

Quite commonly, nondefective retroviruses lacking oncogenes, although unable to transform cells in culture, are capable of inducing a variety of malignancies (426,427). For example, avian leukosis virus, when inoculated into young chickens frequently causes a B-cell lymphoma with a latency of 6 months or more. Close examination of the virus–cell relationship in such tumors reveals a striking result: Virtually all tumors have a provirus inserted in a similar portion of the genome, specifically within the proto-oncogene c-*myc* (158). In the majority of tumors, the *myc*-associated provirus is similarly oriented; it almost always lies within an intron between the first (noncoding) and the second exon, lying in the same transcriptional orientation as well (Fig. 18) The effect of this insertion is to bring the two coding exons of c-*myc* under the transcriptional control of the 3′ LTR. Transcripts encoding

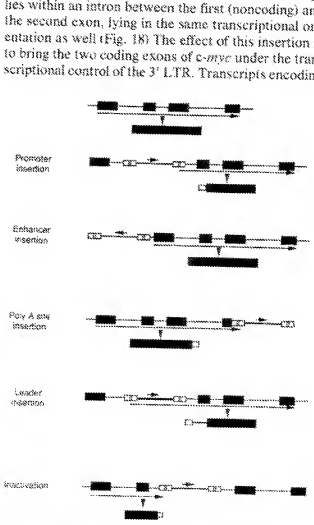

FIG. 18. Effects of retroviral integration. Exons of cellular genes are shown by a solid bar; the primary transcript is shown by an arrow, and the spliced mRNA is shown by a narrow solid bar.

c-myc are found at higher levels than normal and are structurally different in that they contain the R–U5 sequence derived from the LTR at their 5′ ends. Thus, synthesis of these transcripts must be initiated at the promoter sequence within the 3′ LTR. The altered regulation of the *myc* gene in these tumors mimics the case with *v-myc* and, presumably, is a sufficient event to initiate the tumorigenic process in these animals, although other events seem to follow and be necessary for transformation to a fully malignant phenotype (426).

This mechanism of activation of *c-myc* by ALV is usually referred to as "promoter insertion." The use of the 3′ LTR may require some additional sequence rearrangement in the provirus. Most proviruses in the tumors are altered by deletions near, but often not including, the upstream LTR (138,352). These deletions might be important to relieve a block to expression from the downstream LTR.

Insertional activation of a number of proto-oncogenes has been found to be an important feature of pathogenesis by nontransforming retroviruses and has been quite valuable in identifying potential new oncogenes not seen in other ways such as incorporation into retroviruses. Several different mechanisms have been found to be involved (Fig. 18). Interestingly, the promoter insertion mechanism seen in ALV-induced lymphomas is relatively uncommon, perhaps reflecting the necessity of viral mutations as well as the insertion. Other possible mechanisms of gene activation by provirus insertion include: (a) "enhancer insertion," in which insertion of a provirus upstream of a natural or cryptic promoter of a proto-oncogene but in the opposite transcriptional orientation causes deregulated expression from that promoter (426); (b) "leader insertion," where expression of a portion of a proto-oncogene is in the form of multiply spliced readthrough transcripts initiated within the 5′ LTR of a provirus inserted within an intron (302); and (c) "terminator insertion," in which the poly(A) addition signal of the 5′ LTR of a provirus inserted near the 3′ end of a gene results in a truncated mRNA lacking a signal promoting its normally rapid turnover, thereby increasing its effective concentration (308; A. Berns, *personal communication*).

It should be remembered that insertion of a provirus in a position to do this sort of damage must be quite rare on a per-cell basis, but considering the total number of infected cells in the target organ, such an insertion can be statistically certain. Also, in many cases there is good evidence that the initial insertion does not suffice for the full transformation of the target cell to a tumor cell but, rather, that additional mutagenic events must take place, themselves often involving activation by proviral insertion (435,439).

Other Mechanisms

Spleen-Focus-Forming Virus

The several strains of Friend murine leukemia virus induce a variety of diseases of red blood cell precursors, characterized by a very rapid (but nonmalignant) overgrowth of erythroblasts, which often progresses to a malignant, transplantable, erythroleukemia (426,427). While these diseases can be induced (albeit with reduced efficiency) by nondefective virus alone, many independent strains contain a defective component that accelerates the process and gives rise to rapid focal proliferation of (a) erythroblasts in spleens of infected mice and (b) colonies of erythroid cells after infection of bone marrow cell cultures. This component, known as spleen focus-forming virus (SFFV), is replication-defective and encodes a characteristic product known as gp55, which has been shown to be necessary for the "transforming" properties of the virus (21,22,55,470). In contrast to transforming genes of all other known retroviruses, gp55 is not the product of a typical oncogene. Rather, it is a modified *env* gene product, derived by a series of events including (a) acquisition of an altered sequence by recombination with an endogenous provirus *env* gene, (b) a deletion extending from the middle of SU into the TM coding region, and (c) a rearrangement extending the membrane-spanning domain (62). How these alterations combine to convert a retrovirus envelope protein into an oncogene-like protein is not understood. It can be hypothesized that it may (by chance) resemble an erythroid growth factor and thus stimulate growth of infected cells.

HTLV

As discussed in more detail in Reference 476, HTLV-I and related viruses can immortalize lymphoid cells in culture, and infection of cells *in vivo* can lead to their malignant transformation (albeit rarely and with a very long latent period). The mechanism of transformation by these viruses must be different from other retroviruses, since (a) no oncogene is present and (b) consistent proviral integration sites of the type that characterize insertional activation of proto-oncogenes cannot be found. Because expression of the *trans*-activator *tax* can also affect the expression of a number of other genes [in particular, the T-cell growth factor IL-2 and its receptor as well (11,146,221,235,365,382)], it has been hypothesized that this sort of activation might lead to an autocrine mechanism of transformation in which a cell, by secreting its own growth factor, might stimulate itself to divide continuously (266). Direct proof for such a mechanism is lacking, since

expression of *tax* alone is insufficient to immortalize cells in culture. Expression of *tax* in transgenic mice does lead to a benign tumor, but the tumor is very different from the natural disease (299).

Osteopetrosis

This disease can be induced in chickens by certain strains of ALV (426,427). It involves the overgrowth of osteoblasts, particularly in the long bones of the leg, and leads to gross thickening of these bones. It is apparently not a clonal disease and is not a malignancy. The mechanism of its induction is far from clear. Close examination of infected tissue reveals unusually high levels of virus replication and large amounts of unintegrated virus DNA (354), characteristics more commonly associated with cytopathic effects by retroviruses (see below). Perhaps the extensive virus replication results in disruption of the complicated balance between bone growth and dissolution that characterizes the normal tissue. Genetic exchange of fragments between osteopetrotic and lymphoma-inducing strains implicate a region at the beginning of gag as important in conferring the difference (355), but a good mechanistic explanation is not yet forthcoming.

Cytopathic Effects

Although retrovirus infection can, and often does, proceed without noticeable detriment to the infected cell, there are a number of well-characterized virus–cell interactions that lead to its death. The most prominent cytopathic virus–cell interaction is found with HIV and other lentiviruses, discussed extensively in Chapter 53 and References 477 and 478. In addition to these well-known examples, there are a number of models for degenerative disease caused by various oncoviruses closely related to the noncytopathic tumor viruses. These include wasting and immunodeficiency diseases inducible in cats, mice, and monkeys by variants of FeLV and MLV and by a D-type virus, respectively (9,154,159,292). Unfortunately, the temptation to name these conditions FAIDS, MAIDS, and SAIDS was too great to resist. Other variants of MLV have been associated with neurological disease (84, 195,471), and some reticuloendotheliosis virus strains cause degenerative disease (427). At present, there is little insight available into the mechanism by which retrovirus infection leads to death of the host cell. A number of studies have been undertaken to compare the genomes of strains of varying pathology (by exchange of restriction fragments, for example) and to identify gene products important to the difference. In most cases, differences between pathogenic and nonpathogenic variants have been mapped to the *env* gene—in particular, the region encoding the SU protein (195,318,419). In the MAIDS and FAIDS cases, an additional interesting feature of the pathogenic isolates is the repeated isolation of a variety of replication-defective variants of the virus, often carrying large deletions and also harboring some other determinant necessary for disease induction (9,318). Although apparently necessary for the disease process, the precise contribution of these to the pathogenesis is unclear. Defective viruses also arise along with highly cytopathic ALV variants after rapid repeated passage in cell culture, but in this case they do not seem necessary for the cytopathic effect (66,453).

There are several possible mechanisms by which a retrovirus infection might lead to cell destruction: immune system-mediated killing of infected cells; direct toxicity of a virus gene product; and extensive replication of the virus, leading to overwhelming of functions necessary for survival of the cell. All three mechanisms have been put forward as explanations for HIV-associated cytopathology. In cell cultures, certain ALV strains and reticuloendotheliosis viruses exhibit strong cytopathic effects; moreover, useful plaque assays, analogous to those used routinely for other types of viruses, have been developed (90,459). Results using the cell-culture systems are concordant with the hypothesis that cell killing is due to overreplication of virus within the infected cell. Typically, cultures infected with a cytopathic virus have large numbers of copies of unintegrated DNA a few days after infection, indicating that they have been subjected to repeated cycles of virus replication (210). Survivors that outgrow in the culture typically contain only a small number of integrated proviruses. That the appearance of excessive amounts of DNA can be due to complete replication cycles is suggested by the observation that cytopathic ALV variants confer reduced resistance to superinfection (463). Also, detailed mapping studies indicate that the only significant difference between cytopathic and noncytopathic viruses is the ability of the former to use a specific receptor—that for subgroup B virus (90). The conclusion that ensues from these considerations is that cytopathology can be a consequence of the breakdown of the mechanisms that usually act to prevent "replication" of viral DNA within a single infected cell—the combination of post-budding maturation with infection-competent virions, and a very strong and rapidly developing level of resistance to superinfection. Variation in receptor used or in properties of the env protein can apparently decrease the effectiveness of the resistance and permit repeated infection of the cell with the progeny of the provirus it carries, probably leading to (a) an excessive burden of the viral DNA or some other product and (b) consequent death of the cell.

Whether the same mechanism is also important in

killing by HIV remains to be seen. Certainly, very large amounts of unintegrated DNA are often found in infected cells before they die; furthermore, with the correct virus–cell combination, levels of virus protein as high as 40% of the total cell protein have been reported (392). The issue is still open to debate, however. The formation of syncytia as a result of interaction of the env protein on the surface of one cell with CD4 on its neighbor has also been implicated (240,388). So far, mutants have not been very useful in resolving this question. Some relatively noncytopathic variants still display high levels of unintegrated DNA (115), whereas cytopathic interactions can often be seen in the absence of syncytium formation (391).

Insertional Inactivation

The insertion of a provirus into the cell genome is a mutagenic event. Obviously, a provirus inserted into a coding region of a gene will disrupt the function of that gene. To date, no pathogenic process has been conclusively shown to involve this mechanism. This is not too surprising, since (a) most infected cells are diploid and (b) the probability of insertion into the two copies of the same gene at the same time is very small. Also, the occasional loss of gene function in an individual cell usually would go unnoticed unless the event could be amplified, such as by inactivation of a "recessive oncogene." One possible case of this has recently come to light. Progression to true malignancy of tumors induced by Friend MLV often involves proviruses inserted in the gene encoding a protein called p53 (288), discovered as a possible cellular accessory protein binding to transforming products encoded by DNA tumor viruses (see Chapter 13). These results were first interpreted as signaling activation of a cellular proto-oncogene, as in the cases discussed above. However, there is increasing evidence, including the presence in some tumors of a provirus in each allele of p53, that the important event might be inactivation, rather than activation, of its expression (80,163). If so, then p53 might be the second identified example (after Rb) of a gene whose function is in some way necessary to block transformation of cells to a malignant state.

Insertional inactivation of cellular genes has been observed in several cases of endogenous provirus integration (see below) and in several model systems in tissue culture. For example, MLV infection has been used to inactivate expression of *src* from a resident RSV provirus (445), of hypoxanthine phosphoribosyl transferase (HPRT) in mouse teratocarcinoma cells [whose inability to express the virus and therefore superinfection resistance renders them repeatedly infectible (211)], and of β-2 microglobulin on mouse B-cells (118). In all cases, the insertion events were even less frequent than expected, and their detection required both a very powerful selection and that the target gene be haploid. The use of retroviral insertion as an approach to identifying genes encoding a selectable function in cell culture has been repeatedly proposed; to date, however, no new genes have been identified using this sort of "tagging."

Genetics

Retroviruses display a wealth of genetic phenomena that contribute a remarkable level of plasticity to the genome. First, they have a high mutation rate and can, under the appropriate selective conditions, rapidly accumulate both point mutations and major rearrangements of sequence. Second, they display a frequency of genetic recombination unapproached by other systems. Finally, they have the unique ability to acquire foreign sequences, such as oncogenes, and express them as part of their genomes. All of these features together give this virus group the capacity to rapidly evolve into new niches. This sort of evolution can be seen in many settings; in some strains of mice, there is a predictable evolution leading from several benign endogenous proviruses to one capable of causing a tumor and resulting in the premature death of every mouse in the strain (427). In HIV, it results in a high level of sequence and (presumably) antigenic diversity in the *env* gene and its product (63), as well as leading to the appearance of drug-resistant variants in AIDS patients treated with AZT (228) (see Chapter 54). In some ALV and FeLV strains, it leads to the reproducible isolation of a specific oncogene in animals infected with a nontransforming virus (278,291,293,341).

Sequence Variation

Before considering mechanisms of variation, it must be emphasized that the rate of genetic variation is a composite of three variables: the relevant mutation rate per replication cycle, the number of replication cycles per unit time, and the selective advantage or disadvantage possessed by the variant virus. Usually, the third variable is the most important in determining the rate of variation, and it is likely, given the complex combinations of functions encoded within the genome, that no mutation is selectively neutral (66).

Point Mutations

Although much work (discussed below) has been done on the error rate of reverse transcriptases *in vitro*, relatively few studies have addressed the mutation rate (as opposed to the rate of variation) *in vivo*. An ex-

periment in which accumulation of mutations was measured in the course of repeated passages permitted a rough estimate of around 10^{-4} per generation (plus or minus a factor of 3) for RSV (66). A similar estimate (1.4×10^{-4}) was obtained from an experiment utilizing a hybridization technique, in which cloned virus was subjected to rapid recloning and in which the progeny were analyzed for point changes (232). Using a rather different approach—the reversion of a specific point mutation in a dominant selectable gene inserted into a spleen necrosis virus-derived vector—a lower rate of 2×10^{-5} per generation was estimated (93). These results, although obtained with different methods, are reasonably consistent with one another. Considering that there is likely to be considerable variation in mutation rate from one base to the next, the first two approaches are likely to be biased toward the more frequent events. All are within the range of error rates estimated for reserve transcriptase *in vitro*.

These mutation rates are in the range of rates estimated for RNA viruses in general. Similar figures have been obtained for bacteriophage Qβ (16) and for vesicular stomatitis virus (402,403), but up to 10-fold lower rates have been estimated for poliovirus and influenza virus (325). Thus, while mutation rates of retroviruses may be higher than most viruses, they are not truly exceptional.

It is common to refer all blame for mutation to reverse transcriptase. It should, however, be kept in mind that retroviruses are unique in their use of three different enzyme systems for their replication—reverse transcriptase, cell DNA polymerase, and RNA polymerase II. The potential contribution of each of these to mutations will be considered separately, as follows:

Reverse transcriptase. Reverse transcriptase lacks an exonucleolytic (proofreading) activity and seems to have a corresponding high frequency of base misincorporation *in vitro*. Error frequencies ranging over a factor of 100, from 3×10^{-3} to 3×10^{-5} misincorporations per copying, have been reported for the enzyme from various viruses (139,223,251,282,350,386). Since the extremely high rates quoted are based on misincorporation with homopolymeric templates and probably do not closely resemble the natural case, a more reliable estimate would lie within a 10-fold range centered on about one error in 10^4 bases incorporated, a value in rough agreement with measured mutation rates. The reverse transcriptase of HIV seems to be much more error-prone when compared head to head with that of ALSV or MLV (339,349). The significance of this suggestive finding remains to be thoroughly explored. It is important to note that the rate of misincorporation is not likely to be uniform across the whole genome. Rather, it seems to be strongly dependent on the specific context, with some types of errors much more likely than others (349,350).

DNA polymerase. Provirus replication occurs only as a regular part of chromosome replication and must have a similar error rate, one that is likely to be so low as to be negligible for the purposes of this discussion. For example, analysis of a 200-base region of five endogenous proviruses inserted independently into the mouse germline prior to inbreeding, and therefore separated by at least 100 animal generations from one another, revealed no difference in sequence—an error frequency of less than 10^{-5} per animal generation (412). Indeed, an endogenous provirus known to have been inserted into the primate germline more than 5 million years ago still displays considerable sequence similarity to exogenous viruses of mice (345,400).

RNA polymerase. The error rate of RNA polymerase II is not known (and could be very difficult to measure) but is unlikely to be much different from that of other RNA-synthesizing enzymes or from that of reverse transcriptase (i.e., about 10^{-4} plus or minus a factor of 3). If so, then it is worth noting that errors in genomic RNA synthesis may be as important as those in reverse transcription in generating sequence variation. Also noteworthy is that there would be no point for the virus to encode a highly accurate reverse transcriptase if half the replication cycle is carried out by relatively inaccurate pol II.

Rearrangements

Retrovirus genomes are subject to a high (but undetermined) rate of intragenomic rearrangements—deletions, duplications, inversions, or combinations of these. Indeed, a survey of the literature on cloning of viral DNA from infected cells indicates that perhaps half of the DNAs so obtained have suffered some sort of rearrangement (e.g., see ref. 378). In the majority of cases, the rearrangements are so severe as to render the genome inactive, but there are some important exceptions. For example, rearrangements in the LTR of murine leukemia viruses are important determinants of differential pathogenicity (137,380); genetic variation within the *env* gene of different HIV isolates seems to include reduplication of short sequences (63); and specific types of defective mutants such as SFFV (see refs. 62 and 318) can be important pathogens in their own right.

Although the formation of variants has not been analyzed in detail, it is almost certainly a consequence of specific aberrations in reverse transcription. [While rearrangements involving integrated viral DNA can sometimes be detected if the selection is strong enough (e.g., see ref. 6), such events seem to occur at a very

low rate (a few in a million) and cannot account for the vast majority of rearrangements.] Lesions in the process of reverse transcription include mispriming, premature termination of one end or the other, incorrect end-to-end strand transfer, and foldback copying of a newly completed strand to form inversions. The most common type of rearrangement, however, is reduplication or deletion of sequence, most likely due to aberrant "jumping" of reverse transcriptase from one RNA template to another or within a template. These mutants can then be considered to be the consequence of an incorrect recombination event (see next section). It is noteworthy that while deletion and reduplication often involve the use of homologous sequences (178,316), they do not always do so, and there are instances of such rearrangements in the absence of detectable homology or other sequence features (452).

Recombination

One of the most remarkable features of retrovirus genetics is the extraordinarily high rate of recombination. There is probably no other biological system that displays the capacity for exchange of genetic information to the same extent as retroviruses; moreover, all retrovirus systems tested, including HIV (170), undergo recombination at high rates (59). Although most commonly observed between infecting exogenous viruses, recombination can also readily be observed between exogenous and endogeneous viruses and between virus and unrelated host-cell information (as in oncogene capture). Although the frequency of recombination has not been precisely measured, it is so high that in a usual experiment (i.e., coinfecting a cell culture with a mixture of viruses differing in two selectable markers, such as transforming ability and host range, and then selecting a recombination between them), markers as close as 1 kb are found to segregate independently, as if unlinked (59). This free exchange implies that a population of virus containing sequence variations and allowed to interact (by coinfection of the same cells) should be considered to be homogenized across the genome such that all possible combinations of variants are present at any one time. This has important implications for the variation of pathogens such as HIV and for the potential of populations of viruses to generate new combinations of variants which might have very different properties.

Although two plausible models for retrovirus recombination have been proposed (59,199), neither has been subjected to a critical test. It is clear that recombination only occurs following infection with heterozygous virions (i.e., those produced by cells coinfected with two parental viruses) (243). If a cell is simply doubly infected, recombinants are not observed in that

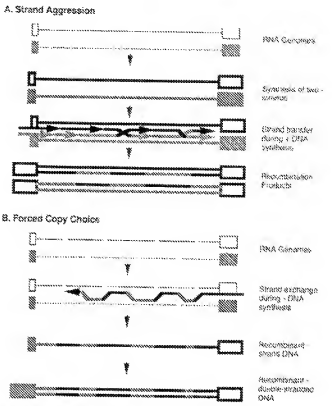

FIG. 19. Possible mechanisms of retrovirus recombination. In both cases, DNA synthesis beginning with a virus with a heterozygous genome leads to the formation of recombinants. In the first case (200), two full-length negative strands are synthesized, and recombination occurs by strand exchange during positive-strand synthesis. In the second case (59), "jumps" by reverse transcriptase wherever a break is encountered lead to synthesis of a recombinant negative strand.

cell. From this requirement, it follows that recombination must be an early event, occurring prior to integration, presumably related to the process of viral DNA synthesis. [It is important to note that the introduction of transfection techniques of naked retroviral DNA (as with all other DNAs) into cells leads to high rates of recombination (14), but this process—although useful for mapping—has nothing to do with the natural case.]

The two models proposed are illustrated in Fig. 19. The first model (199) proposes that exchange of a newly made fragmentary positive strand copied from one genome onto a negative-strand copy of the other could lead to heteroduplex proviruses that would segregate wild-type and recombinant virus after integration and cell division. While concordant with electron-microscopic observation of the products of *in vitro* reverse transcriptase reactions (198), this scheme requires that both members of the diploid genome serve as templates simultaneously, and it does not readily

explain recombination involving large regions of nonhomology, such as around oncogenes.

The second model (59) is based on two observations: (i) Reverse transcriptase efficiently switches templates when it encounters an end while still elongating a growing chain, and (ii) retrovirus genomes can have considerable numbers of breaks and yet remain associated in an intact complex. Thus, when reverse transcriptase encounters such a break, it can behave exactly as it would at the end of a genome—that is, switch to the homologous sequence on the other genome and continue synthesis. The product of this reaction would be a negative strand containing a mosaic of information from both genomes (depending on the number of preexisting breaks). This mechanism not only would promote high-frequency exchange of information but would have the important additional benefit of repairing preexisting RNA breakage and relieving the virus of using extraordinary measures to shield its genome from the ravages of the extracellular environment. Since reverse transcriptase can make strand transfers in the absence of extended homology (indicating that affinity of enzyme for template may be a stronger driving force for jumping than for base-pairing), a similar mechanism can be readily invoked to explain other types of genomic rearrangements, as well as capture of cellular sequences.

Oncogene Capture

Retroviruses are the only group of animal viruses that acquire cellular genes and convert them to oncogenes, so the mechanism of oncogene acquisition must be connected with special features of retroviruses. An attractive model has been developed from the original observation of Goldfarb and Weinberg (137), who studied a deletion mutant of Harvey murine sarcoma virus (Ha-MSV) which contained the H-*ras* oncogene but lacked any homology to MLV at its 3′ end. They found that virus heterozygous for this truncated genome and for a wild-type MuLV genome would yield transforming virus with good frequency, indicating that the 3′ end of the genome could be readily repaired by illegitimate recombination. Since such partial genomes closely resemble those that might be derived from aberrancies surrounding the integration event, the following model was proposed (418) (Fig. 20):

1. The provirus integrates upstream from a proto-oncogene. As discussed above, integrations of this sort can also lead directly to oncogene "activation."
2. A deletion at the DNA level joins the proto-oncogene to a portion of the provirus.
3. The joint viral-oncogene transcript is processed (by splicing of the oncogene portion, for example) and incorporated into heterozygous virions along with a wild-type genome (from another provirus in the same cell).
4. An illegitimate recombination event during reverse transcription restores the 3′ end of the viral genome.

Note that the final structure of the oncogene containing virus genome may also be determined by additional rearrangements, as well as by point mutations within the oncogene itself, so that the structure observed in viruses need not exactly reflect the initial events. A variation of this model, based on the observation that correct formation of viral 3′ ends (by cleavage and polyadenylation) is an inefficient process, has also been proposed (Fig. 20) (162), differing from the other as follows: After integration adjacent to an oncogene, a joint readthrough transcript (perhaps processed by splicing in the oncogene portion) is incorporated into heterozygous virus, along with a wild-type virus genome, and then a pair of illegitimate recombination events during reverse transcription incorporates the oncogene into a viral genome.

Endogenous Viruses

A particularly noteworthy and unique feature of retroviruses is their presence as inherited elements in the germline of many (perhaps all) vertebrates, where they behave as reasonably stable Mendelian genes (60,411). It has been estimated that as much as 5–10% of the mammalian genome may consist of elements introduced by mechanisms involving reverse transcription (430). Perhaps 10% of these are identifiable on structural grounds as provirus-like; that is, they contain (a) LTRs flanked by short direct repeats and (b) primer-binding sites that flank internal coding regions with detectable relationship to *gag* and *pol*. The remainder comprise a variety of "retroelements," including movable elements such as the long interspersed repeat sequences (or LINES) whose motion is apparently directed by machinery encoded within the element, as well as families of sequences such as processed pseudogenes and the abundant Alu elements in human DNA which appear to have been moved by processes that include reverse transcription but which do not themselves encode such activity (13,430). For the most part, this discussion will be confined to the provirus-like elements—in particular, the endogenous proviruses (i.e., those related to known viruses). The remainder of provirus-like retrotransposable elements—such as copia and similar elements in flies (288), Ty1 and other yeast transposons (30), and some plant elements (52,105,142,457)—seem to use mechanisms quite similar to those of retroviruses for their motion and genetic association with the host. They differ principally in the apparent lack of an extracellular or virion

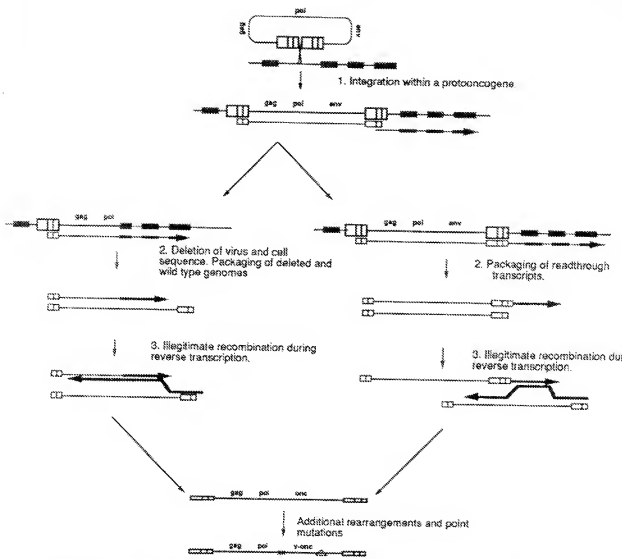

FIG. 20. Possible mechanisms for oncogene capture by retroviruses. The scheme on the left (418) differs from that on the right (162) by the requirement for a specific deletion in DNA of the infected cell which fuses viral and cellular sequences. As shown on the right, this fusion can also be accomplished by reverse transcriptase acting on readthrough transcripts. In either case, important additional mutations and rearrangements probably occur during subsequent virus replication.

phase. Indeed, if many of the endogenous proviruses now known were first identified solely by their appearance in the genome without reference to known viruses, they too would be classed as retrotransposons.

Distribution

Active endogenous proviruses (those closely related to infectious viruses and potentially encoding them) are sporadically distributed among species. When present, they are unevenly distributed among individuals within a species. They have clearly become associated with the germline subsequent to speciation and are continuing to show reinsertion into it in modern times. Chickens, for example, contain from one to about four or so proviruses closely related to ALSV and residing at a variety of sites in the genome, whereas none have been detected in related species such as turkeys or quail or even other members of the same genus (60). In humans, no active proviruses have been detected, although large numbers of more distantly related (but probably noninfectious) elements can be found (152,173,231,261,315); moreover, infectious endogenous viruses have been detected in other primates (423).

In mice, four classes of elements totalling 500–1,000 members have been identified as endogenous proviruses, although two of them are not known to encode

virus genomes. The VL30 (for virus-like 30S) proviruses encode RNAs which apparently are not competent but which are efficiently encapsidated in MLV virions. They may be an ancient element devolved to the status of a passenger virus, only capable of replicating by using endogenous or exogenous MLV as its helper. The intracisternal A particles (IAPs) are often expressed as large numbers of intracellular particles with active reverse transcriptase, but they have never been seen in virions. The two endogenous proviruses that can encode infectious virus are the B-type, closely related to exogenous MMTV, and the C-type, related to exogenous MLV. The former is present in 0–4 copies per mouse; the latter is present in about 50–60 (60,411). The C-type proviruses can be further subdivided into four classes according to properties of the envelope gene (and other linked sequence characteristics): ecotropic, xenotropic, polytropic, and modified polytropic (412), with the latter three being about 10-fold more abundant (and somewhat more ancient) than the ecotropic class (169,219,332). Similar (but less well studied) groups of endogenous proviruses are also present in cats.

Detailed analysis of the distribution of the C-type proviruses in mice, including assignment of over 100 different proviruses in different inbred strains to specific chromosomal locations (119; W. Frankel, J. P. Stoye, B. Taylor, and J. M. Coffin, *manuscript in preparation*), has revealed the following general principles: (a) Proviruses are stably present at many locations on all chromosomes. (b) No two inbred strains are alike in their composition of proviruses, and only a few are found in all strains. (c) Different proviruses within a group are very closely related to one another, differing principally by their chromosomal location, and seem to be separated by only a few viral replication cycles. (d) In most cases, acquisition of new germline proviruses is quite rare and seems to correlate with the extent of virus replication in the animal, implying that any apparent "transposition" seen is a result of virus replication. An interesting exception to the rarity of such events is seen in crosses involving SWR mice in which offspring of viremic mothers acquire new germline proviruses at an extraordinarily high rate (17,250,397), making a potentially valuable system for insertional mutagenesis. (e) Similarly, the rate of proviral loss is quite low. In several different types of experiments, values of about 4.5×10^{-6} to 10^{-7} have been estimated (374). The most common mechanism of loss seems to be recombination between the two LTRs, leaving a solo LTR behind. No evidence whatsoever has ever been presented for specific excision or intracellular transposition or any other mechanism of movement of proviruses except through the normal replication pathway.

The overall relationship between the active endogenous proviruses and the host thus appears to be a sort of rough steady state in which a slow rate of acquisition of new proviruses approximately balances the random losses. These are not the only retrovirus-related information present, however. If individuals who lack active proviruses (humans or quail, for example) are tested for more distantly related sequences by nucleic acid hybridization techniques, it is usually easy to find many such elements (96,411). These are clearly proviruses, but they differ from the active ones as follows: (a) They exhibit more general species distribution; (b) there is similarity of location among different individuals within a species; and (c) they are grossly defective, usually containing many mutations that would prevent their replication. Often, the two LTRs have diverged somewhat in sequence (345); this is a sure sign of long residence in the germline, as is the presence of the same elements in the same location in humans and chimpanzees (400). All these properties are consistent with these proviruses being the degraded residue of viruses active in the distant past, some of whose proviruses, although defective, still remain in the germline. These elements are clearly defective proviruses, not some other sort of element. They must be derived from ancient periods of retrovirus infection colonizing the germline and represent groups of endogenous viruses within which all active members have died out, leaving only the residue of defective proviruses remaining in the genome. Thus, retroviruses are the only virus group to have left a true fossil record—the endogenous proviruses present in all of us. Study of these could be quite rewarding for the light they can shed on evolutionary processes.

Properties of Endogenous Viruses

Endogenous proviruses are not simply germline copies of exogenous infectious agents. Rather, they have a number of special features, some of which seem to be specific adaptations to the endogenous lifestyle, others of which are apparently enforced by residence in the host germline:

1. Endogenous proviruses are usually transcriptionally silent. This effect is an epigenetic one. If expression of a nondefective endogenous provirus is induced and infects new cells, it is often capable of directing at least moderate levels of transcription. The most likely explanation for this effect involves suppression of expression by extensive methylation at CpG dinucleotides (60,411), and possibly other mechanisms (such as heritable modifications in chromatin structure) as well. Agents, such as 5-azacytidine, that reverse DNA methylation can efficiently induce expression, at least transiently, of endogenous proviruses. The silencing of these proviruses is not a feature

unique to them: Proviruses of exogenous viruses, like Moloney MLV, are also affected the same way after introduction into the germline of mice (188,189,411). Presumably, this reflects a host mechanism for dealing with foreign DNA. Often, the transcriptional repression is very strong but not absolute. Under conditions where the endogenous provirus is capable of replicating in the host animal, even a very low level of expression can provide sufficient virus to cause a generalized infection. Some animals (such as AKR mice) usually become viremic at an early age (60,427). Not all endogenous proviruses are strongly transcriptionally suppressed. In chickens, for example, some defective proviruses are expressed at levels 100-fold higher than others (60). This phenomenon may be due to position effects of adjoining chromatin structure on the provirus (69,147). In some mice, coordinate expression of a number of C-type proviruses at different locations is controlled by a single genetic locus (designated *Gv-1*) (237,238).

2. Endogenous proviruses are often defective, typically differing from the canonical wild-type virus by deletions or point mutations that render them incapable of yielding infectious virus. One endogenous MLV provirus, for example, contains a mutation near the beginning of *gag* that prevents myristylation (71). Another has an AUA codon instead of AUG at the beginning of *env*, preventing initiation of translation (412).

3. Many endogenous viruses, although replication-competent, are often unable to replicate in the animals in which they are found, due to the mismatch in receptor ("xenotropism") discussed above.

4. With one exception, viruses induced from endogenous proviruses are nonpathogenic and are oftentimes less efficient in replication than their exogenous relatives. This difference presumably represents an adaptation to the endogenous lifestyle; viruses of significant pathogenicity would reduce the reproductive potential of their host and be counterselected. The difference in replicative rate is largely attributable to sequence differences in the LTR, as shown by genetic exchange studies with both MLV and ALV (46,171,434). In particular, endogenous provirus LTRs have significantly reduced enhancer activity as compared to their pathogenic cousins. Other sequence differences affecting pathogenicity but not replication rate have also been observed (43,46,353) but not explained.

Effects on the Host

Beneficial Effects

When endogenous retroviruses were first described, it became popular to propose ways in which they or related elements might provide some essential service to the host organism. The variability in these elements, along with the breeding of individuals lacking certain types of proviruses, laid these ideas to rest, at least in their simplest form. It is more useful to consider endogenous proviruses as a type of selfish element, most likely derived from exogenous infectious viruses but specifically adapted to residence in the germline. It is possible that the presence of some endogenous proviruses confers a beneficial effect relevant to pathogenesis by exogenous viruses. Expression of env proteins, for example, can induce interference toward infection by viruses that use the same receptor or can induce tolerance to related antigens that these proteins might encode, thus minimizing pathogenic effects due to immunological killing of infected cells (76).

Pathogenic Effects

In general, endogenous viruses, when expressed, do not exert pathogenic effects on their host. Chickens viremic with the endogenous virus, known as RAV-O, are indistinguishable from normal birds. The only endogenous virus known to directly cause disease is MMTV, in which inheritance of an endogenous provirus gives rise to a significant incidence of mammary carcinoma, even in mice foster-nursed on virus-free strains to eliminate maternal transmission (60).

A more complex example was created when some strains of mice were inbred to select for high incidence of spontaneous leukemia (60,411,427). Mice of strains such as AKR and C58 typically die at about 1 year of age with a T-cell lymphoma induced by a retrovirus derived from proviruses endogenous to the particular strains. While the ultimate molecular event in oncogenesis involves insertional activation of one or more oncogenes by new proviruses (373,450), and possibly other events as well (81), a complex set of rearrangements is involved in the generation of the ultimate oncogenic viruses themselves (Fig. 21). First, an ecotropic provirus is expressed and replicates widely in the animal around the time of birth. Second, this virus recombines with a xenotropic provirus to provide a new LTR. Third, a specific rearrangement occurs in the LTR sequence to duplicate the enhancer region. Fourth, recombination with a polytropic provirus alters the envelope SU (gp70) coding region and gives its host range to the virus. Finally, the new virus, known as MCF for its ability to induce foci on mink cells, infects target T cells, and a set of events including integration of proviral DNA adjacent to a proto-oncogene, as well as other less well understood steps (81), leads to transformation into a tumor cell.

These specific events occur synchronously (as can be detected using specific DNA probes) in virtually

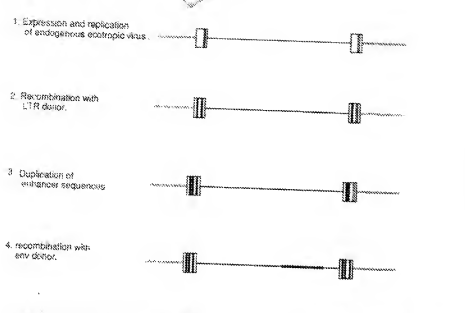

FIG. 21. Virological events during lymphomagenesis in AKR mice. It is important to keep in mind that each of the events is relatively rare, but the specific products are selected by their increased replication ability or by their effects on the target cell.

every mouse of affected strains. There is no imaginable basis by which the specific recombinations and rearrangements can be directed to occur; rather, the products seen most are those selected out of a very large pool of randomly occurring events.

The selective forces involved are not well understood. The LTR alterations presumably act to enhance the ability of the virus to replicate in the target tissue. The *env* alteration provides the recombinant virus with the ability to use a different receptor (343), but the role of this difference is less than obvious. In all of this genetic contortion, it is striking that the only genomic variation seen is that which is selected for, although initial recombination events must have generated a very large number of different combinations. Comparison of nucleotide sequence of recombinant and parental viruses reveals very little random sequence variation (47,172,224). This example points up the critical role of selective forces in shaping retroviral genomes.

Mutagenic Effects

The insertion of an endogenous provirus into a new location on the genome is a potentially mutagenic event. Considering the number and variety of endogenous proviruses in some species such as mice, it would not be surprising to find that some are associated with known mutations, and genetic analyses have supplied several examples. Two well-known mutations of mice, *d* (dilute brown) and *hr* (hairless), have each been found to be genetically inseparable from specific endogenous proviruses in standard genetic experiments. In both cases, reversion of the mutation was accompanied by loss of the provirus, leaving behind a solo LTR (70,180,193,413) and providing proof of a causal relationship. In both cases, the provirus has been used as a tag, from which flanking DNA corresponding to part of the gene could be cloned and analyzed. Considering that there are well over 100 different endogenous C-type proviruses known, this type of analysis could prove quite valuable in identifying DNA sequences corresponding to genes not easily accessed by other methods. Indeed, based on linkage analysis, it has been estimated that as many as 5% of recessive mutations in the mouse may be associated with the insertion of endogenous C-type proviruses (119,413). In a related example, introduction of exogenous proviruses into the mouse germline (by infection of early embryos) has also been found to be mutagenic. For example, one such provirus (known as Mov-13) was found to cause death of the homozygous embryo in a mid-embryonic stage, due to its insertion into a gene whose product becomes essential to the survival of the embryo at that point—the gene encoding α-1 collagen (155).

Most of the mutations caused by insertion of endogenous proviruses are recessive, resulting from loss of expression of the affected gene. In principle, dominant effects, analogous to oncogene activation, might also be expected but are rarely observed, probably reflecting the lack of transcriptional activity of proviruses passed through the germline. Although no active endogenous viruses are known to have activated expression of a host gene, one interesting example of such activation by a more ancient provirus has been reported. In this example, expression of a duplicated

copy of the mouse C4 complement gene (*Slp*, whose product is known as sex-limited protein) is activated and made responsive to stimulation by androgen (254) as a result of the presence upstream of an ancient, deleted C-type provirus whose LTR provides the necessary hormone-responsive enhancer (399).

Other Effects

Even aside from their direct role as mutagens, the number, variety, and antiquity of endogenous proviruses imply that they have played important roles in shaping the vertebrate genome throughout evolution (13,430). It is possible that they contribute directly not only to their own expression and reinsertion but also to that of unrelated sequence as well. Some defective endogenous retrovirus-related elements such as VL30 RNA are efficiently packaged into infectious virions by proteins of the unrelated MLV (2,60,62,184,273), and at least one small RNA related to repeated elements in the genome is copackaged into mLV virions and is reverse-transcribed (54). Although it is believed that processed pseudogenes and middle repetitive sequences move by processes involving reverse transcription, the source of the reverse transcriptase is not obvious, since none has ever been detected in uninfected cells. There is now some evidence from *in vitro* models that the necessary reverse transcriptase might be provided by retrovirus infection. In at least two cases, the rare transfer to a recipient cell of a cDNA copy of mRNA carrying a selectable marker has been observed after expression of the marker in cells expressing genome-free retrovirus virions (89,242). Although such transfers were rare, and the structure of the new DNA was not perfectly concordant with precedent established by the pseudogenes, these systems provide precedent for, and a model of, possible retrovirus-assisted motion of the much larger groups of cell elements.

REFERENCES

1. Adam MA, Miller AD. Identification of a signal in a murine retrovirus that is sufficient for packaging of nonretroviral RNA into virions. *J Virol* 1988;62:3802–3806.
2. Adams SE, Rattlen PD, Stanway CA, et al. Complete nucleotide sequence of a murine VL30 retro-element. *Mol Cell Biol* 1988;8:2989–2998.
3. Ahmad N, Venkatesan S. Nef protein of HIV-1 is a transcriptional repressor of HIV-1 LTR. *Science* 1988;241:1481–1485.
4. Albritton LM, Tseng L, Scadden D, Cunningham JM. A putative murine ecotropic retrovirus receptor gene encodes a multiple membrane-spanning protein and confers susceptibility to virus infection. *Cell* 1989;57:659–666.
5. Aldovini A, DeRossi A, Feinberg MB, Wong-Staal F, Franchini G. Molecular analysis of a deletion mutant provirus of type I human T-cell lymphotropic virus: evidence for a doubly spliced x-lor mRNA. *Proc Natl Acad Sci USA* 1986;83:38–43.
6. Alexander F, Leis J, Soltis DA, et al. Proteolytic processing of avian sarcoma and leukosis viruses pol-endo recombinant proteins reveals another *pol* gene domain. *J Virol* 1987;61:534–542.
7. Arrigo S, Beemon K. Regulation of Rous sarcoma virus RNA splicing and stability. *Mol Cell Biol* 1988;18:4858–4867.
8. Arrigo S, Yun M, Beemon K. *Cis*-acting regulatory elements within *gag* genes of avian retroviruses. *Mol Cell Biol* 1987;7:388–392.
9. Arya SK. 3′-*orf* and *sor* genes of human immunodeficiency virus: *In vitro* transcription-translation and immunoreactive domains. *Proc Natl Acad Sci USA* 1987;84:5429–5433.
10. Aziz DC, Hanna Z, Jolicoeur P. Severe immunodeficiency disease induced by a defective murine leukaemia virus. *Nature* 1989;338:505–508.
11. Ball JK, Diggelmann H, Dekaban GA, et al. Alterations in the U3 region of the long terminal repeat of an infectious thymotropic type B retrovirus. *J Virol* 1988;62:2985–2993.
12. Ballard DW, Bohnlein E, Lowenthal JW, Wano Y, Franza BR, Greene WC. HTLV-I tax induces cellular proteins that activate κB element in the IL-2 receptor α gene. *Science* 1988;241:1652–1655.
13. Baltimore D. RNA-dependent DNA polymerase in virions of RNA tumour viruses. *Nature* 1970;226:1209–1211.
14. Baltimore D. Retroviruses and retrotransposons: the role of reverse transcription in shaping the eukaryotic genome. *Cell* 1985;40:481–482.
15. Bandyopadhyay PK, Watanabe S, Temin HM. Recombination of transfected DNA's in vertebrate cell culture. *Proc Natl Acad Sci* 1984;81:3476–3480.
16. Barklis E, Mulligan RC, Jaenisch R. Chromosomal position or virus mutation permits retrovirus expression in embryonal carcinoma cells. *Cell* 1986;47:391–399.
17. Batschelet E, Domingo E, Weissmann C. The proportion of revertant and mutant phage in a growing population, as a function of mutation and growth rate. *Gene* 1976;1:27–32.
18. Bautch VL. Genetic background affects integration frequency of ecotropic proviral sequences into the mouse germ line. *J Virol* 1986;60:693–701.
19. Bedinger P, Moriarty A, von Borstel II RC, Donovan NJ, Steimer KS, Littman DR. Internalization of the human immunodeficiency virus does not require the cytoplasmic domain of CD4. *Nature* 1988;334:162–165.
20. Bender MA, Palmer TD, Gelinas RE, Miller AD. Evidence that the packaging signal of Moloney murine leukemia virus extends into the gag region. *J Virol* 1987;61:1639–1646.
21. Berg J. Potential metal-binding domains in nucleic acid binding proteins. *Science* 1986;232:485–487.
22. Berger SA, Sanderson N, Bernstein A, Rankins WD. Induction of the early stages of Friend erythroleukemia with helper-free Friend spleen focus-forming virus. *Proc Natl Acad Sci USA* 1985;82:6913–6917.
23. Bestwick RK, Rankins WD, Kabat D. Roles of helper and defective retroviral genomes in murine erythroleukemia: studies of spleen focus-forming virus in the absence of helper. *J Virol* 1985;56:660–664.
24. Biegalke B, Linial M. Retention or loss of v-mil sequences after propagation of MH2 virus in vivo or in vitro. *J Virol* 1987;61:1949–1956.
25. Birnstiel ML, Busslinger M, Strub K. Transcription termination and 3′ processing: the end is in site! *Cell* 1985;41:349–359.
26. Bishop JM. Clues to the puzzle of purpose. *Nature* 1985;316:483–484.
27. Bishop JM, Varmus HE. Functions and origins of retroviral transforming genes. In: Weiss R, Teich N, Varmus H, Coffin J, eds. *RNA tumor viruses*. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory, 1982;999–1108.
28. Bishop JM, Varmus HE. Functions and origins of retroviral transforming genes. In: Weiss R, Teich N, Varmus H, Coffin J, eds. *RNA tumor viruses, 2nd edition*. Cold Spring Harbor New York: Cold Spring Harbor Laboratory, 1985;249–356.
29. Bister K, Jansen HW. Oncogenes in retroviruses and cells: biochemistry and molecular genetics. *Adv Cancer Res* 1986;47:99–188.
30. Bodor J, Svoboda J. The LTR, v-src, LTR provirus generated

in the mammalian genome by src mRNA reverse transcription and integration. *J Virol* 1989;63:1015–1018.
30. Boeke JD. Retrotransposons. In: Domingo E, Holland JJ, Ahlquist P, eds. *RNA genetics, vol II: retroviruses, viroids, and RNA recombination*. Boca Raton, Florida: CRC Press, 1988;59–103.
31. Bohnlein S, Hauber J, Cullen BR. Identification of a U5-specific sequence required for efficient polyadenylation within the human immunodeficiency virus long terminal repeat. *J Virol* 1988;63:421–424.
32. Bohnlein E, Siekevitz M, Ballard DW, et al. Stimulation of the human immunodeficiency virus type 1 enhancer by the human T-cell leukemia virus type I tax gene product involves the action of inducible cellular proteins. *J Virol* 1989;63:1578–1586.
33. Bolognesi DP, Montelaro RC, Frank H, Schafer W. Assembly of type C oncornaviruses: a model. *Science* 1978;199:183–186.
34. Boone LR, Glover PL, Innes CL, Niver LA, Bondurant MC, Yang WK, Fv-1 N- and B-tropism-specific sequences in murine leukemia virus and related endogenous proviral genomes. *J Virol* 1988;62:2644–2650.
35. Boral AL, Okenquist SA, Lenz J. Identification of the SL3-3 virus enhancer core as a T-lymphoma cell-specific element. *J Virol* 1988;63:76–84.
36. Bosze Z, Thiesen H, Charnay P. A transcriptional enhancer with specificity for erythroid cells is located in the long terminal repeat of the Friend murine leukemia virus. *EMBO J* 1986;5:1615–1624.
37. Bova CA, Manfredi JP, Swanstrom R. *env* genes of avian retroviruses: nucleotide sequence and molecular recombinants define host range determinants. *Virology* 1986;152:343–354.
38. Bova CA, Olsen JC, Swanstrom R. The avian retrovirus *env* gene family: molecular analysis of host range and antigenic variants. *J Virol* 1988;62:75–83.
39. Bowerman B, Brown PO, Bishop JM, Varmus HE. A nucleoprotein complex mediates the integration of retroviral DNA. *Genes Dev* 1989;3:469–478.
40. Brady J, Jeang K, Duvall J, Khoury G. Identification of p40x-responsive regulatory sequences within the human T-cell leukemia virus type I long terminal repeat. *J Virol* 1987;61:2175–2181.
41. Brierley I, Boursnell MEG, Binns MM, et al. An efficient ribosomal frameshifting signal in the polymerase-encoding region of the coronavirus IBV. *EMBO J* 1987;6:3779–3787.
42. Broome S, Gilbert W. Rous sarcoma virus encodes a transcriptional activator. *Cell* 1985;40:537–546.
43. Brown DW, Blair BP, Robinson HL. Long terminal repeat (LTR) sequences, env, and a region near the 5′ LTR influence the pathogenic potential of recombinants between Rous-associated virus types 0 and 1. *J Virol* 1988;62:3431–3437.
44. Brown PO, Bowerman B, Varmus HE, Bishop JM. Retroviral integration: structure of the initial covalent product and its precursor, and a role for the IN protein. *Proc Natl Acad Sci USA* 1989;86:2525–2529.
45. Brown PO, Bowerman B, Varmus HE, Bishop JM. Correct integration of retroviral DNA *in vitro*. *Cell* 1987;49:347–356.
46. Brown DW, Robinson HL. Role of *RAV-0* genes in the permissive replication of subgroup E avian leukosis viruses on host 15B cv1 CEF. *Virology* 1988;162:239–242.
47. Buller RS, Ahmed A, Portis JL. Identification of two forms of an endogenous murine retroviral *env* gene linked to the *Rmcf* locus. *J Virol* 1987;61:29–34.
48. Carlberg K, Beemon K. Proposed *gag*-encoded transcriptional activator is not necessary for Rous sarcoma virus replication or transformation. *J Virol* 1988;62:4003–4008.
49. Carlberg K, Ryden TA, Beemon K. Localization and footprinting of an enhancer within the avian sarcoma virus *gag* gene. *J Virol* 1988;62:1617–1624.
50. Celander D, Haseltine WA. Glucocorticoid regulation of murine leukemia virus transcription element is specified by determinants within the viral enhancer region. *J Virol* 1987;61:269–275.
51. Celander D, Isu BL, Haseltine WA. Regulatory elements within the murine leukemia virus enhancer regions mediate glucocorticoid responsiveness. *J Virol* 1988;62:1314–1322.
52. Chandler VL, Wilbot V. DNA modification of a maize transposable element correlates with loss of activity. *Proc Natl Acad Sci USA* 1986;83:1767–1771.
53. Chen ISY. Regulation of AIDS virus expression. *Cell* 1986;47:1–2.
54. Chen P, Cywinski A, Taylor JM. Reverse transcription of 7S L RNA by an avian retrovirus. *J Virol* 1985;54:278–284.
55. Chang S, Wolff L, Ruscetti S. Sequences responsible for the altered erythropoietin responsiveness in spleen focus-forming virus strain SFFVp-infected cells are localized to a 678-base-pair region at the 3′ end of the envelope gene. *J Virol* 1987;61:1651–1664.
56. Clayton LK, Hussey RE, Steinbrich R, Ramachandran H, Hussain Y, Reinherz EL. Substitution of murine for human CD4 residues identifies amino acids critical for HIV-gp120 binding. *Nature* 1988;335:363–366.
57. Cobrinik D, Katz R, Terry R, Skalka AM, Leis J. Avian sarcoma and leukosis virus pol-endonuclease recognition of the tandem long terminal repeat junction: minimum site required for cleavage is also required for viral growth. *J Virol* 1987;61:1999–2008.
58. Cobrinik D, Soskey L, Leis J. A retroviral RNA secondary structure required for efficient initiation of reverse transcription. *J Virol* 1988;62:3622–3630.
59. Coffin JM. Structure, replication, and recombination of retrovirus genomes: some unifying hypotheses. *J Gen Virol* 1979;42:1–26.
60. Coffin JM. Endogenous viruses. In: Weiss R, Teich N, Varmus H, Coffin J, eds. *RNA tumor viruses*. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory, 1982;1109–1204.
61. Coffin JM. Structure of the retroviral genome. In: Weiss R, Teich N, Varmus H, Coffin J, eds. *RNA tumor viruses*. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory, 1982:261–369.
62. Coffin JM. Genome structure. In: *RNA tumor viruses. 2nd edition*. Weiss R, Teich N, Varmus H, Coffin J, eds. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory. 1985: 17–74.
63. Coffin JM. Genetic variation in AIDS viruses. *Cell* 1986;46:1–4.
64. Coffin JM. Replication of retrovirus genomes. In: Domingo E, Holland JJ, Ahlquist P, eds. *RNA genetics, vol II: retroviruses, viroids, and RNA recombination*. Boca Raton, Florida: CRC Press, 1988;3–22.
65. Coffin JM, Haseltine WA. Terminal redundancy and the origin of replication of Rous sarcoma virus RNA. *Proc Natl Acad Sci USA* 1977;74:1908–1912.
66. Coffin JM, Tsichlis PN, Barker CS, Voynow S. Variation in avian retrovirus genomes. *Ann NY Acad Sci* 1980;354:410–425.
67. Colicelli J, Goff SP. Isolation of a recombinant murine leukemia virus utilizing a new primer tRNA. *J Virol* 1986;57:37–45.
68. Colicelli J, Goff SP. Sequence and spacing requirements of a retrovirus integration site. *J Mol Biol* 1988;199:47–59.
69. Conklin KF, Groudine M. Varied interactions between proviruses and adjacent host chromatin. *Mol Cell Biol* 1986;6:3999–4007.
70. Copeland NG, Hutchinson KW, Jenkins NA. Excision of the DBA ecotropic provirus in dilute coat-color revertants of mice occurs by homologous recombination involving the viral LTRs. *Cell* 1983;33:379–387.
71. Copeland NG, Jenkins NA, Nexo B, et al. Poorly expressed endogenous ecotropic provirus of DBA/2 mice encodes a mutant Pr65gag protein that is not myristylated. *J Virol* 1988;62:479–487.
72. Cordingley MG, Riegel AT, Hager GL. Steroid-dependent interaction of transcription factors with the inducible promoter of mouse mammary tumor virus *in vivo*. *Cell* 1987;48:261–270.
73. Craigen WJ, Caskey CT. Translational frameshifting: where will it stop? *Cell* 1987;50:1–2.
74. Craigie R, Mizuuchi K. Transposition of the Mu DNA: joining of Mu to target DNA can be uncoupled from cleavage at the ends of Mu. *Cell* 1987;51:493–501.
75. Crawford S, Goff SP. Mutations in gag proteins p12 and p15 of Moloney murine leukemia virus block early stages of infection. *J Virol* 1984;49:909–917.
76. Crittenden LB, McMahon S, Halpern MS, Fadly AM. Embry-

one infection with the endogenous avian leukosis virus Rous-associated virus-0 alters responses to exogenous avian leukosis virus infection. *J Virol* 1987;61:722–725.

77. Cullen BR, Lomedico PT, Ju G. Transcriptional interference in avian retroviruses: implications for the promoter insertion model of leukemogenesis. *Nature* 1984;307:241–244.
78. Deleted in proof.
79. Dalgleish AG, Beverly PCL, Clapham PR, Crawford DH, Greaves MF, Weiss RA. The CD4 (T4) antigen is an essential component of the receptor for the AIDS retrovirus. *Nature* 1984;312:763–767.
80. David YB, Prideaux VR, Chow V, Benchimol S, Bernstein A. Inactivation of the p53 oncogene by internal deletion of retroviral integration in erythroleukemic cell lines induced by Friend leukemia virus. *Oncogene* 1988;3:179–186.
81. Davis BR, Brightman BK, Chandy KG, Fan H. Characterization of a preleukemic gene induced by Moloney murine leukemia virus: evidence for two infection events during leukemogenesis. *Proc Natl Acad Sci USA* 1987;84:4875–4879.
82. Davis MG, Kenney SC, Kamine J, Pagano JS, Huang E. Immediate-early gene region of human cytomegalovirus *trans*-activates the promoter of human immunodeficiency virus. *Proc Natl Acad Sci USA* 1987;84:8642–8646.
83. Davis NL, Rueckert RR. Properties of a ribonucleoprotein particle isolated from Nonidet P-40-treated Rous sarcoma virus. *J Virol* 1972;10:1010–1020.
84. DesGroseillers L, Rassart E, Robitaille Y, Jolicoeur P. Retrovirus-induced spongiform encephalopathy: the 3′-end long terminal repeat-containing viral sequences influence the incidence of the disease and the specificity of the neurological syndrome. *Proc Natl Acad Sci USA* 1985;82:8818–8822.
85. Dickson C, Eisenman R, Fan H. Protein synthesis and assembly. In: Weiss R, Teich N, Varmus H, Coffin J, eds. *RNA tumor viruses, 2nd edition*. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory, 1985;135–146.
86. Dickson C, Eisenmann R, Fan H, Hunter E, Teich N. Protein biosynthesis and assembly. In: Weiss RA, Teich N, Varmus HE, Coffin JM, *RNA tumor viruses*. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory, 1982;513–648.
87. Donehower LA. Analysis of mutant Moloney murine leukemia viruses containing linker insertion mutations in the 3′ region of pol. *J Virol* 1988;62:3958–3964.
88. Doolittle RF, Feng D, Johnson MS, McClure MA. Origins and evolutionary relationships of retroviruses. *Q Rev Biol* 1989;64:1–30.
89. Duesberg R, Temin HM. Retroviral vector system for the study of cDNA gene formation. *Mol Cell Biol* 1988;8:2328–2334.
90. Dorner AJ, Coffin JM. Determinants for receptor interaction and cell killing on the avian retrovirus glycoprotein gp85. *Cell* 1986;45:365–374.
91. Dorner AJ, Stoye JP, Coffin JM. Molecular basis of host range variation in avian retroviruses. *J Virol* 1985;53:32–39.
92. Dougherty JP, Temin HM. A promoterless retroviral vector indicates that there are sequences in U3 required for 3′ processing. *Proc Natl Acad Sci USA* 1987;84:1197–1201.
93. Dougherty JP, Temin, H.M. High mutation rate of a spleen necrosis virus-based retrovirus vector. *Mol Cell Biol* 1986;6:4387–4395.
94. Dudley JP. Mouse mammary tumor proviruses from a T-cell lymphoma are associated with the retroposon LIMd. *J Virol* 1988;62:472–478.
95. Dunwiddie CT, Resnick R, Boyce-Jacino M, Alegre JN, Faras AJ. Molecular cloning and characterization of gag-, pol-, env-related gene sequences in the ev– chicken. *J Virol* 1986;59:669–675.
96. Dutta A, Dorai T, Hamaguchi H. The putative *trans*-activator in the MAgag region of Rous sarcoma virus is not required for cell transformation. *J Virol* 1988;62:4767–4769.
97. Duyk G, Longiaru M, Cornish D, et al. Circles with two tandem long terminal repeats are specifically cleaved by *pol* gene-associated endonuclease from avian sarcoma and leukosis viruses: nucleotide sequences required for site-specific cleavage. *J Virol* 1985;56:589–599.
98. Eichinger DJ, Boeke JD. The DNA intermediate in yeast Ty1 element transposition copurifies with virus-like particles: free Ty1 transposition. *Cell* 1988;54:955–966.
99. Enfield D, Hunter E. Oligomeric structure of a prototype retrovirus glycoprotein. *Proc Natl Acad Sci USA* 1988;85:8688–8692.
100. Ellis J, Bernstein A. Gene targeting with retroviral vectors: recombination by gene conversion into regions of nonhomology. *Mol Cell Biol* 1989;9:1621–1627.
101. Embretson JE, Temin HM. Lack of competition results in efficient packaging of heterologous murine retroviral RNAs and reticuloendotheliosis virus encapsidation minus RNAs by the reticuloendotheliosis virus helper cell line. *J Virol* 1987;61:2675–2683.
102. Fan H, Chute H, Chao E, Pattengale PK. Leukemogenicity of Moloney murine leukemia viruses carrying polyoma enhancer sequences in the long terminal repeat is dependent on the nature of the inserted polyoma sequences. *Virology* 1988;166:58–65.
103. Fan H, Mittal S, Chute H, Chao E, Pattengale PK. Rearrangements and insertions in the Moloney murine leukemia virus long terminal repeat alter biological properties *in vivo* and *in vitro*. *J Virol* 1986;60:204–214.
104. Farmerie WG, Loeb DD, Casavant NC, Hutchinson III CA, Edgell MH, Swanstrom R. Expression and processing of the AIDS virus reverse transcriptase in *Escherichia coli*. *Science* 1987;236:305–308.
105. Fedoroff NV. About maize transposable elements and development. *Cell* 1989;56:181–191.
106. Feinberg MB, Jarrett RF, Aldovini A, Gallo RC, Wong-Staal F. HTLV-III expression and production involve complex regulation at the levels of splicing and translation of viral RNA. *Cell* 1986;46:807–817.
107. Felber BK, Hadzopoulou-Cladaras M, Cladaras C, Copeland T, Pavlakis G. rev protein of human immunodeficiency virus 1 affects the stability and transport of the viral mRNA. *Proc Natl Acad Sci USA* 1989;86:1495–1499.
108. Felber BK, Paskalis H, Kleinman-Ewing C, The pX protein of HTLV-I is a transcriptional activator of its long terminal repeats. *Science* 1985;229:675–679.
109. Feisenstein KM, Goff SP. Expression of the gag-pol fusion protein of Moloney murine leukemia virus without gag protein does not include virion formation or proteolytic processing. *J Virol* 1988;62:2179–2182.
110. Feng Y, Hatfield DL, Ren A, Levin JG. Translational readthrough of the murine leukemia virus gag gene amber codon does not require virus-induced alteration of tRNA. *J Virol* 1989;63:2405–2410.
111. Feng S, Holland EC. HIV-1 tat *trans*-activation requires the loop sequence within tar. *Nature* 1988;334:165–167.
112. Fauser G, Taketo M, Haneeak RC, Fan H. Two blocks in Moloney murine leukemia virus expression in undifferentiated F9 embryonal carcinoma cells as determined by transient expression assays. *J Virol* 1989;63:2317–2324.
113. Flinton WI, Champoux JJ. RNA-primed initiation of Moloney murine leukemia virus plus strands by reverse transcriptase *in vitro*. *J Virol* 1984;51:26–33.
114. Fisher AG, Ensoli B, Ivanoff L, et al. The *sor* gene of HIV-1 is required for efficient virus transmission *in vitro*. *Science* 1987;237:888–893.
115. Fisher AG, Ratner L, Mitsuya H, et al. Infectious mutants of HTLV-III with changes in the 3′ region and markedly reduced cytopathic effects. *Science* 1986;233:655–659.
116. Flügel RM, Rethwilm A, Maurer B, Darai G. Nucleotide sequence of the *env* gene and its flanking regions of the human spumaretrovirus reveals two novel genes. *EMBO J* 1987;6:2077–2084.
117. Franchini G, Robert-Guroff M, Ghrayeb J, Chang NT, Wong-Staal F. Cytoplasmic localization of the HTLV-III 3′ orf protein in cultured T cells. *Virology* 1986;155:593–599.
118. Frankel W, Potter TA, Rosenberg N, Lenz J, Rapan TV. Retroviral insertional mutagenesis of a target allele in a heterozygous murine cell line. *Proc Natl Acad Sci USA* 1985;82:6600–6604.
119. Frankel WN, Stoye JP, Taylor BA, Coffin JM. Genetic analysis of endogenous xenotropic murine leukemia viruses: association with two common mouse mutations and the viral restriction locus *Fv-1*. *J Virol* 1989;63:1763–1774.

120. Franz T, Hilberg F, Seliger B, Stocking C, Ostertag W. Retroviral mutants efficiently expressed in embryonal carcinoma cells. *Proc Natl Acad Sci USA* 1986;83:3292–3296.
121. Fu X, Katz RA, Skuntz AM, Leis J. Site-directed mutagenesis of the avian retrovirus nucleocapsid protein pp12. mutation which affects RNA binding *in vitro* blocks viral replication. *J Biol Chem* 1988;263:2134–2139.
122. Fu S, Phillips N, Jentoft J, Tuazon PT, Traugh JA, Leis J. Site-specific phosphorylation of avian retrovirus nucleocapsid protein pp12 regulates its binding to RNA. *J Biol Chem* 1985;260:9941–9947.
123. Fuetterer J, Hohn T. Involvement of enolecapsids in reverse transcription: a general phenomenon? *TIBS* 1987;12:92–95.
124. Fujisawa J, Seiki M, Sato M, Yoshida M. A transcriptional enhancer sequence of HTLV-I is responsible for *trans*-activation mediated by pilb, of HTLV-I. *EMBO J* 1986;5:713–718.
125. Fujiwara T, Mizuuchi K. Retroviral DNA integration: structure of an integration intermediate. *Cell* 1988;54:497–504.
126. Gallaher WR. Detection of a fusion peptide sequence in the transmembrane protein of human immunodeficiency virus. *Cell* 1987;50:327–328.
127. Gallis B, Linial M, Eisenmann R. An avian oncovirus mutant deficient in genomic RNA: characterization of the packaged RNA as cellular messenger RNA. *Virology* 1979;94:146–161.
128. Gallo R, Wong-Staal F, Montagnier L, Haseltine WA, Yoshida M. HIV/HTLV gene nomenclature. *Nature* 1988;333:504.
129. Garfinkel DJ, Boeke JD, Fink GR. Ty element transposition: reverse transcriptase and virus-like particles. *Cell* 1985;42:507–517.
130. Gautsch J, Wilson MC. Delayed *de novo* methylation in teratocarcinoma suggests additional tissue-specific mechanisms for controlling gene expression. *Nature* 1983;301:32–37.
131. Gelinas C, Temin HM. Nondefective spleen necrosis virus-derived vectors define the upper size limit for packaging reticuloendotheliosus viruses. *Proc Natl Acad Sci USA* 1986;83:9211–9215.
132. Gendelman HE, Phelps W, Feigenbaum L, et al. *Trans*-activation of the human immunodeficiency virus long terminal repeat sequence by DNA viruses. *Proc Natl Acad Sci USA* 1986;83:9759–9763.
133. Gilboa E, Mitra SW, Goff S, Baltimore D. A detailed model of reverse transcription and tests of crucial aspects. *Cell* 1979;18:93–100.
134. Gimble JM, Duh E, Ostrove JM, Gendelman HE, Max EE, Rabson AB. Activation of the human immunodeficiency virus long terminal repeat by herpes simplex virus type 1 is associated with induction of a nuclear factor that binds to the NF-κB/core enhancer sequence. *J Virol* 1988;62:4104–4112.
135. Goff SP. The genetics of murine leukemia viruses. *Curr Top Microbiol Immunol* 1984;112:45–69.
136. Goldfarb MP, Weinberg RA. Generation of novel, biologically active Harvey sarcoma viruses via apparent illegitimate recombination. *J Virol* 1981;38:136–150.
137. Golemis E, Li Y, Fredrickson TN, Hartley JW, Hopkins N. Distinct segments within the enhancer region collaborate to specify the type of leukemia induced by nondefective Friend and Moloney virus. *J Virol* 1988;63:328–337.
138. Grodeenow MM, Hayward WS. 5′ long terminal repeats of myc-associated proviruses appear structurally intact but are functionally impaired in tumors induced by avian leukosis viruses. *J Virol* 1987;61:2486–2498.
139. Gopinathan KP, Weymouth LA, Kunkel TA, Loeb, LA. Mutagenesis *in vitro* by DNA polymerase from an RNA tumor virus. *Nature* 1979;278:857–859.
140. Gorelick RJ, Henderson LE, Hanser JP, Rein A. Point mutants of Moloney murine leukemia virus that fail to package viral RNA: evidence for specific RNA recognition by a "zinc finger-like" protein sequence. *Proc Natl Acad Sci USA* 1988;85:8420–8424.
141. Gorman CM, Rigby PWJ, Lane DP. Negative regulation of viral enhancers in undifferentiated embryonic stem cells. *Cell* 1985;42:519–526.
142. Grandbastien M, Spielmann A, Caboche M. Tnt1, a mobile retroviral-like transposable element of tobacco isolated by plant cell genetics. *Nature* 1989;337:376–380.
143. Grandgenett DP, Vora AC. Site-specific nicking at the avian retrovirus LTR circle junction by the viral pp32 DNA endonuclease. *Nucleic Acids Res* 1985;13:6205–6221.
144. Grandgenett DP, Vora AC, Swanstrom R, Olsen JC. Nuclease mechanism of the avian retrovirus pp32 endonuclease. *J Virol* 1986;58:970–974.
145. Graves BJ, Johnson PF, McKnight SL. Homologous recognition of a promoter domain common to the MSV LTR and the HSV tk gene. *Cell* 1986;44:565–576.
146. Greene WC, Leonard WJ, Wano Y, et al. *Trans*-activator gene of HTLV-II induces IL-2 receptor and IL-2 cellular gene expression. *Science* 1986;232:877–881.
147. Groudine M, Conkin KE. Chromatin structure and do novo methylation of sperm DNA: implications for activation of the paternal genome. *Science* 1985;228:1061–1068.
148. Hackett PB, Petersen RB, Hensel CH, et al. Synthesis *in vitro* of a seven amino acid peptide encoded in the leader RNA of Rous sarcoma virus. *J Mol Biol* 1986;190:45–58.
149. Hadzopoulou-Cladaras M, Felber BK, Cladaras C, Athanassopoulos A, Tse A, Pavlakis GN. The rev (trs/art) protein of human immunodeficiency virus type 1 affects viral mRNA and protein expression via a *cis*-acting sequence in the env region. *J Virol* 1989;63:1265–1274.
150. Hagino-Yamagishi K, Donehower LA, Varmus HE. Retroviral DNA integrated during infection by an integration-deficient mutant of murine leukemia virus is oligomeric. *J Virol* 1987;61:1964–1971.
151. Hanecak R, Pattengale PK, Fan H. Addition or substitution of simian virus 40 enhancer sequences into the Moloney murine leukemia virus (M-MuLV) long terminal repeat yields infectious M-MuLV with altered biological properties. *J Virol* 1988;62:2427–2436.
152. Harada F, Tsukada N, Kato N. Isolation of three kinds of human endogenous retrovirus-like sequences using $(RNA^{Pro}$ as a probe. *Nucleic Acids Res* 1987;15:9153–9162.
153. Harris JD, Blum H, Scott J, Traynor B, Ventura P, Haase A. Slow virus visna: reproduction *in vitro* of virus from extrachromosomal DNA. *Proc Natl Acad Sci USA* 1984;81:7212–7215.
154. Hartley JW, Frederickson TN, Yetter RA, Makino M, Morse HC. Retrovirus-induced murine acquired immunodeficiency syndrome: natural history of infection and differing susceptibility of inbred mouse strains. *J Virol* 1989;63:1223–1231.
155. Hartung S, Jaenisch R, Breindl M. Retrovirus insertion inactivates mouse α1(I) collagen gene by blocking initiation of transcription. *Nature* 1986;320:365–367.
156. Haseltine WA, Wong-Staal F. The molecular biology of the AIDS virus. *Sci Am* 1988,259:52–62.
157. Hauber J, Cullen BR. Mutational analysis of the *trans*-activation responsive region of the human immunodeficiency virus I long terminal repeat. *J Virol* 1988;62:673–679.
158. Hayward WS, Neel BG, Astrin SM. Activation of a cellular *onc* gene by promoter insertion in ALV-induced lymphomas. *Nature* 1981;290:475–480.
159. Heidecker G, Lerche MW, Lowenstine LJ, et al. Induction of simian acquired immune deficiency syndrome (SAIDS) with a molecular clone of a type D SAIDS retrovirus. *J Virol* 1987;61:3066–3071.
160. Henderson LE, Sowder RC, Copeland TD, Benveniste RE, Oroszlan S. Isolation and characterization of a novel protein (X-ORF product) from SIV and HIV-2. *Science* 1988;241:199–201.
161. Herman SA, Coffin JM. Differential transcription from the long terminal repeats of integrated avian leukosis virus DNA. *J Virol* 1986;60:497–505.
162. Herman SA, Coffin JM. Efficient packaging of readthrough RNA in ALV: implications for oncogene transduction. *Science* 1987;236:845–848.
163. Hicks G, Mowat M. Integration of Friend murine leukemia virus into both alleles of the p53 oncogene in an erythroleukemic cell line. *J Virol* 1988;62:4752–4755.
164. Hilberg F, Stocking C, Ostertag W, Grez M. Functional analysis of a retroviral host-range mutant: altered long terminal repeat sequences allow expression in embryonal carcinoma cells. *Proc Natl Acad Sci USA* 1987;84:5232–5236.
165. Hippenmeyer PJ, Grandgenett DP. Requirement of the avian retrovirus pp32 DNA binding protein domain for replication. *Virology* 1984;137:358–370.

166. Hirano A, Wong, T. Functional interaction between transcriptional elements in the long terminal repeat of reticuloendotheliosis virus: cooperative DNA binding of promoter- and enhancer-specific factors. *Mol Cell Biol* 1988;8:5232–5244
167. Hirsch V, Reidel N, Mullins JI. The genome organization of STLV-3 is similar to that of the AIDS virus except for a truncated transmembrane protein. *Cell* 1987;49:307–319.
168. Hiti A, Henderson LE, Copeland TD, Sowder RC, Hixson CV, Oroszlan S. Characterization of murine mammary tumor virus *gag-pro* gene products and ribosomal frameshift site by protein sequencing. *Proc Natl Acad Sci USA* 1987;84:7041–7045.
169. Hoggan MD, O'Neill RR, Kozak CA. Nonecotropic murine leukemia viruses in BALB/c and NFS/N mice; characterization of the BALB/c Rtv-1 provirus and the single NFS endogenous xenotrope. *J Virol* 1986;60:980–985.
170. Hoggan MD, Willey RL, Strebel K, Martin MA, Repaske R. Genetic recombination of human immunodeficiency virus. *J Virol* 1989;63:1455–1459.
171. Holland CA, Thomas CY, Chattopadhyay SK, Kochne C, O'Donnell PV. Influence of enhancer sequences on thymotropism and leukemogenicity of mink cell focus-forming viruses. *J Virol* 1989;63:1284–1292.
172. Holland CA, Wozney J, Hopkins N. Nucleotide sequence of the *spl9* gene of murine retrovirus MCF 247. *J Virol* 1983;47:413–420.
173. Horn TM, Huebner K, Croce C, Callahan R. Chromosomal locations of members of a family of novel endogenous human retroviral genomes. *J Virol* 1986;58:955–959.
174. Horvat RT, Wood C, Balachandran N. Transactivation of human immunodeficiency virus promoter by human herpesvirus 6. *J Virol* 1989;63:970–973.
175. Hoxie JA, Alpers JD, Rackowski JL, et al. Alterations in T4 (CD4) protein and mRNA synthesis in cells infected with HIV. *Science* 1986;234:1123–1127.
176. Hsu CL, Fabritius C, Dudley J. Mouse mammary tumor virus proviruses in T-cell lymphomas lack a negative regulatory element in the long terminal repeat. *J Virol* 1988;62:4644–4652.
177. Hsu HW, Schwartzberg P, Goff S. Point mutations in the p30 domain of Moloney murine leukemia virus. *Virology* 1985;142:211–214.
178. Hughes SH, Kosick E. Mutagenesis of the region between env and src of the SR-A strain of Rous sarcoma virus for the purpose of constructing helper independent vectors. *Virology* 1984;136:89–99.
179. Hunter E, Hill E, Hardwick M, Bhown A, Schwartz DE, Tizard R. Complete sequence of the Rous sarcoma virus *env* gene: identification of structural and functional regions of its products *J Virol* 1983;46:920–936.
180. Hutchinson KW, Copeland NG, Jenkins NA. Dilute-coat-color locus of mice: nucleotide sequence analysis of the d + 2J and d + Ha revertant alleles. *Mol Cell Biol* 1986;4:2899–2904.
181. Hwang JV, Gilboa E. Expression of genes introduced into cells by retroviral infection is more efficient than that of genes introduced into cells by DNA transfection. *J Virol* 1984;50:417–424.
182. Inoue J, Yoshida M, Seiki M. Transcriptional (p40x) and post-transcriptional (p27x-III) regulators are required for the expression and replication of the leukemia virus type 1 genes. *Proc Natl Acad Sci USA* 1987;84:3363–3367.
183. Inouye S, Hsu M, Eagle S, Inouye M. Reverse transcriptase associated with the biosynthesis of the branched RNA-linked msDNA in *Myxococcus xanthus*. *Cell* 1989;56:709–717.
184. Iba A, Keshet E. A novel retrovirus-like family in mouse DNA. *J Virol* 1986;59:301–307.
185. Jacks T, Madhani HD, Maslarz FR, Varmus HE. Signals for ribosomal frameshifting events are required for synthesis of mouse mammary tumor virus gag-related polyproteins. *Proc Natl Acad Sci USA* 1987;84:4298–4302.
187. Jacks T, Varmus HE. Expression of the Rous sarcoma virus pol gene by ribosomal frameshifting. *Science* 1985;230:1237–1242.
188. Jahner D, Jaenisch R. Chromosomal position and specific demethylation in enhancer sequences of germ line-transmitted retroviral genomes during mouse development. *Mol Cell Biol* 1985;5:2212–2220.
189. Jahner D, Jaenisch R. Retrovirus-induced *de novo* methylation of flanking host sequences correlates with gene inactivity. *Nature* 1985;315:594–597.
190. Jakobovits A, Smith DH, Jakobovits EB, Capon DJ. A discrete element 3′ of human immunodeficiency virus 1 (HIV-1) and HIV-2 mRNA initiation sites mediates transcriptional activation by an HIV *trans* activator. *Mol Cell Biol* 1988;8:2555–2561.
191. Jameson BA, Rao PE, Kong LI, et al. Location and chemical synthesis of a binding site for HIV-1 on the CD4 protein. *Science* 1988;240:1335–1339.
192. Jeang K, Boros I, Brady J, Radonovich M, Khoury G. Characterization of cellular factors that interact with the human T-cell leukemia virus type I p40x-responsive 21-base-pair sequence. *J Virol* 1988;62:4499–4509.
193. Jenkins NA, Copeland NG, Taylor BA, Lee BK. Dilute (*d*) coat colour mutation of DBA/2J mice is associated with the site of integration of an ecotropic MuLV genome. *Nature* 1981;293:370–374.
194. Johnson MS, McClure MA, Feng D, Gray J, Doolittle RF. Computer analysis of retroviral *pol* genes: assignment of enzymatic functions to specific sequences and homologies with nonviral enzymes. *Proc Natl Acad Sci USA* 1986;83:7648–7652.
195. Jolicoeur P, DesGroseillers L. Nonreciprocal Cas-Br-E murine leukemia virus harbors several determinants of leukemogenicity mapping in different regions of the genome *J Virol* 1985;56:639–643.
196. Jones KA, Kadonaga JT, Luciw PA, Tjian R. Activation of the AIDS retrovirus promoter by the cellular transcription factor. *Science* 1986;232:755–759.
197. Jorgensen EC, Kjeldgaard NO, Pedersen FS, Jorgensen P. A nucleotide substitution in the gag N-terminus of the endogenous ecotropic DBA/2 virus prevents PrPr$_{gag}$ myristylation and virus replicaton. *J Virol* 1988;62:3217–3223.
198. Junghans RP, Boone LR, Skalka AM. Products of reverse transcription in avian retrovirus analyzed by electron microscopy. *J Virol* 1982;43:544–554
199. Junghans RP, Boone LR, Skalka AM. Retroviral DNA H structures: displacement–assimilation model of recombination. *Cell* 1982;30:53–62.
200. Kane SE, Beemon K. Precise localization of m6A in Rous sarcoma virus RNA reveals clustering of methylation sites: implications for RNA processing. *Mol Cell Biol* 1985;5:2298–2306.
201. Kao SY, Calman AF, Luciw PA, Peterlin BM. Antitermination of transcription within the long terminal repeat of HIV-1 by *tat* gene product. *Nature* 1987;330:489–493.
202. Karnitz L, Poon D, Weil A, Chalifey R. Purification and properties of the Rous sarcoma virus internal enhancer binding protein. *Mol Cell Biol* 1989;9:1929–1939.
203. Katoh I, Ikawa Y, Yoshinaka Y. Retrovirus protease characterized by a dimeric aspartic proteinase. *J Virol* 1989;63:2226–2232.
204. Katoh I, Yoshinaka Y, Rein A, Shibuya M, Odaka T, Oroszlan S. Murine leukemia virus maturation: protease region required for conversion from "immature" to "mature" core form and for virus infectivity. *Virology* 1985;145:280–292.
205. Katz RA, Kotler M, Skalka AM. *cis*-acting intron mutations that affect the efficiency of avian retroviral RNA splicing: implications for mechanisms of control. *J Virol* 1988;62:2686–2695.
206. Katz RA, Cullen BR, Malavarca R, Skalka AM. Role of the avian retrovirus mRNA leader in expression: evidence for novel translational control. *Mol Cell Biol* 1986;6:372–379
207. Katz RA, Skalka AM. A C-terminal domain in the avian sarcoma-leukosis virus *pol* gene product is not essential for viral replication *J Virol* 1988;62:528–531.
208. Katz RA, Terry RW, Skalka AM. A conserved cis-acting sequence in the 5′ leader of avian sarcoma virus RNA is required for packaging. *J Virol* 1986;59:163–167.
209. Kearney S, Kamme J, Matkovitz D, Fenrick R, Pagano J. An Epstein–Barr virus immediate-early gene product *trans*-activates expression from the human immunodeficiency virus lon terminal repeat. *Proc Natl Acad Sci* 1988;85:1652–1656.

210. Keshet E, Temin HM. Cell killing by spleen necrosis virus is correlated with a transient accumulation of spleen necrosis virus DNA. *J Virol* 1979;31:376–388.

211. King W, Patel MD, Lobel LI, Goff SP, Nguyen-Huu MC. Insertion mutagenesis of embryonal carcinoma cells by retroviruses. *Science* 1985;228:554–558.

212. Kitano H, Chen IS, Shah NP, Cann AJ, Shimotohno K, Fan H. U3 sequences from HTLV-I and -II LTRs confer pX protein response to a murine leukemia virus LTR. *Science* 1987;235:901–904.

213. Klatzmann D, Champagne E, Chamaret S, et al. T-lymphocyte T4 molecule behaves as the receptor for human retrovirus LAV. *Nature* 1984;312:767–768.

214. Knuth RJ, Hippenmeyer PJ, Misra TK, Grandgenett DP, Muller UR, Fitch WM. Avian retrovirus pp32 DNA binding protein. Preferential binding to the promoter region of long terminal repeat DNA. *Biochemistry* 1986;25:359–359.

215. Kohl NE, Emini EA, Schleif WA, et al. Active human immunodeficiency virus protease is required for viral infectivity. *Proc Natl Acad Sci USA* 1988;85:4686–4690.

216. Kotler M, Katz RA, Danho W, Leis J, Skalka AM. (1988) Synthetic peptides as substrates and inhibitors of a retroviral protease. *Proc Natl Acad Sci USA* 1988;85:4185–4189.

217. Kotler M, Katz RA, Skalka AM. Activity of avian retroviral protease expressed in *Escherichia coli*. *J Virol* 1988;62:2696–2700.

218. Kowalski M, Potz J, Basiripour L, et al. Functional regions of the envelope glycoprotein of human immunodeficiency virus type 1. *Science* 1987;237:1351–1355.

219. Kozak CA, O'Neill RR. Diverse wild mouse origins of xenotropic, mink-cell focus-forming, and two types of ecotropic proviral genes. *J Virol* 1987;61:3082–3088.

220. Kramer RA, Schaber MD, Skalka AM, Ganguly K, Wong-Staal F, Reddy EP. HTLV-III gag protein is processed in yeast cells by the virus pol-protease. *Science* 1986;231:1580–1583.

221. Kronke M, Leonard WJ, Depper JM, Greene WC. Deregulation of interleukin-2 receptor gene expression in HTLV-I-induced adult T-cell leukemia. *Science* 1985;228:1215–1217.

222. Kuchino Y, Beier H, Akita N, Nishimura S. Natural UAG suppressor glutamine tRNA is elevated in mouse cells infected with Moloney murine leukemia virus. *Proc Natl Acad Sci USA* 1987;84:2668–2672.

223. Kunkel TA, Schaaper RM, Loeb LA. Depurination-induced infidelity of deoxynucleic acid synthesis with purified deoxyribonucleic acid replication proteins *in vitro*. *Biochemistry* 1983;22:2378–2384.

224. Laigret F, Repaske R, Boniface K, Rahson AB, Khan AS. Potential progenitor sequences of mink cell focus-forming (MCF) murine leukemia viruses: ecotropic, xenotropic, and MCF-related viral RNAs are detected concurrently in thymus tissues of AKR mice. *J Virol* 1986;62:376–386.

225. Lamins LA, Tschlis PN, Khoury G. Multiple enhancer domains in the 3' terminus of the Prague strain of Rous sarcoma virus. *Nucleic Acids Res* 1984;12:6427–6442.

226. Lampson BC, Sun J, Hsu M, Vallejo-Ramirez J, Inouye S, Inouye M. Reverse transcriptase in a clinical strain of *Escherichia coli*: production of branched RNA-linked msDNA. *Science* 1989;243:1033–1038.

227. Landau NR, Warton M, Littman DR. The envelope glycoprotein on the human immunodeficiency virus binds to the immunoglobulin-like domain of CD4. *Nature* 1988;334:159–162.

228. Larder BA, Darby G, Richman DD. HIV with reduced sensitivity to zidovudine (AZT) isolated during prolonged therapy. *Science* 1989;243:1731–1734.

229. Lasky LA, Nakamura G, Smith DH, et al. Delineation of a region of the human immunodeficiency virus type 1 gp120 glycoprotein critical for interaction with the CD4 receptor. *Cell* 1987;50:975–985.

230. Lee WT, Prakash O, Klein D, Sarkar NH. Structural alterations in the long terminal repeat of an acquired mouse mammary tumor virus provirus in a T-cell leukemia of DBA/2 mice. *Virology* 1987;159:39–48.

231. Leib-Mosch C, Brack B, Werner T, Erfle V, Hehlmann R. Isolation of an SSAV-related endogenous sequence from human DNA. *Virology* 1986;155:666–677.

232. Leider JM, Palese P, Smith FI. Determination of the mutation rate of a retrovirus. *J Virol* 1988;62:3084–3091.

233. Leis J, Baltimore D, Bishop JM, et al. Standardized and simplified nomenclature for proteins common to all retroviruses. *J Virol* 1988;62:1808–1809.

234. Leonard JM, Abramczyk JW, Pazen DS, et al. Development of disease and virus recovery in transgenic mice containing HIV proviral DNA. *Science* 1988;242:1665–1672.

235. Leung K, Nabel GJ. HTLV-I transactivator induces interleukin-2 receptor expression through an NF-kB-like factor. *Nature* 1988;333:776–778.

236. Levantis P, Gillespie DAF, Hart K, Bissell MJ, Wyke JA. Control of expression of an integrated Rous sarcoma provirus in rat cells: role of 5' genomic duplications reveals unexpected patterns of gene transcription and its regulation. *J Virol* 1986;57:907–916.

237. Levy DE, Lerner RA, Wilson MC. The *Gv-1* locus coordinately regulates the expression of multiple endogenous murine retroviruses. *Cell* 41:289–299

238. Levy DE, Lerner RA, Wilson MC. Normal expression of polymorphic endogenous retroviral RNA containing segments identical to mink cell focus-forming virus. *J Virol* 1985;56:691–700.

239. Li Y, Golemis E, Hartley JW, Hopkins N. Disease specificity of nondefective Friend and Moloney murine leukemia viruses is controlled by a small number of nucleotides. *J Virol* 1987;61:693–700.

240. Lifson JD, Reyes GR, McGrath MS, Stein BS, Engelman EG. AIDS retrovirus induced cytopathology: giant cell formation and involvement of CD4 antigen. *Science* 1986;232:1123–1127.

241. Lim D, Maas WK. Reverse transcriptase-dependent synthesis of a covalently linked, branched DNA–RNA compound in *E. coli* B. *Cell* 1989;56:891–904.

242. Linial M. Creation of a processed pseudogene by retroviral infection. *Cell* 1987;49:93–102.

243. Linial M, Blair D. Genetics of retroviruses. In: Weiss R, Teich N, Varmus H, Coffin J, eds. *RNA Tumor Viruses*. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory, 1982;649–783.

244. Linial M, Medeiros E, Hayward WS. An avian oncovirus mutant (SE 21Q1b) deficient in genomic RNA: biological and biochemical characterization. *Cell* 1978;15:1371–1381.

245. Linial M, Miller AD. Retroviral RNA packaging: sequence requirements and implications. *Curr Top Microbiol Immunol* 1989;(*in press*).

246. Linney E, Davis B, Overhauser J, Chao E, Fan H. Nonfunction of a Moloney murine lenkemia virus regulatory sequence in F9 embryonal carcinoma cells. *Nature* 1984;308:470–472.

247. Linney E, Neill SD, Prestridge DS. Retroviral vector gene expression in F9 embryonal carcinoma cells. *J Virol* 1987;61:3248–3253.

248. Lobel LI, Goff SP. Reverse transcription of retroviral genomes: mutations in the terminal repeat sequences. *J Virol* 1985;53:447–455

249. Lobel LI, Patel M, King W, Nguyen-Huu MC, Goff SP. Construction and recovery of viable retroviral genomes carrying a bacterial suppressor transfer RNA gene. *Science* 1985;228:329–332.

250. Lock LF, Keshet E, Gilbert DJ, Jenkins NA, Copeland NG. Studies of the mechanism of spontaneous germline ecotropic provirus acquisition in mice. *EMBO J* 1988;7:4169–4177.

251. Loeb DD, Padgett RW, Hardies SC, et al. The sequence of a large L1Md element reveals a tandemly repeated 5' end and several features found in retrotransposons. *Mol Cell Biol* 1986;6:168–182.

252. Loh TP, Sievert LL, Scott RW. Proviral sequences that restrict retroviral expression in mouse embryonal carcinoma cells. *Mol Cell Biol* 1987;7:3775–3784.

253. Loh TP, Sievert LL, Scott RW. Negative regulation of retrovirus expression in embryonal carcinoma cells mediated by an intragenic domain. *J Virol* 1988;62:4086–4095.

254. Lorem F, Stavenhagen J, Kalff M, Robins DM. A complex androgen-responsive enhancer resides 2 kilobases upstream of the mouse *Slp* gene. *Mol Cell Biol* 1988;8:3250–3260.

255. LoSardo JE, Cupelli LA, Short MK, Berman JW, Lenz J, Dif-

ferences in activities of murine retroviral long terminal repeats in cytotoxic T lymphocytes and T-lymphoma cells. *J Virol* 1989;63:1087–1094.

256. Luk K, Gilmore TD, Panganiban AT. The spleen necrosis virus *int* gene product expressed in *Escherichia coli* has DNA binding activity and mediates *att* and U5-specific DNA multimer formation *in vitro*. *Virology* 1987;157:127–136.

257. Maddon PJ, Dalgleish AG, McDougal JS, Clapham PR, Weiss RA, Axel R. The *T4* gene encodes the AIDS virus receptor and is expressed in the immune system and the brain. *Cell* 1986;47:333–348.

258. Maddon PJ, McDougal JS, Clapham PR, et al. HIV infection does not require endocytosis of its receptor, CD4. *Cell* 1988;54:865–874.

259. Maddon PJ, Molineaux SM, Maddon DE, et al. Structure and expression of the human and mouse *T4* genes. *Proc Natl Acad Sci USA* 1987;84:9155–9159.

260. Madore N, Panet A, Honigman A. Translation of *gag*, *pro*, and *pol* gene products of human T-cell leukemia virus type 2. *J Virol* 63:2606–2604.

261. Mager DX, Greeman JD. Human endogenous retrovirus-like genome with Type C pol sequences and gag sequences related to human t-cell lymphotropic viruses. *J Virol* 1987;61:4060–4066.

262. Malim MH, Hauber J, Le S, Maizel JV, Cullen BR. The HIV-1 rev *trans*-activator acts through a structured target sequence to activate nuclear export of unspliced viral mRNA. *Nature* 1989;338:254–257.

263. Malim MH, Hauber J, Fenrick R, Cullen BR. Immunodeficiency virus rev *trans*-activator modulates the expression of the viral regulatory genes. *Nature* 1988;335:181–183.

264. Mann R, Baltimore D. Varying the position of a retrovirus packaging sequence results in the encapsidation of both unspliced and spliced RNAs. *J Virol* 1985;54:401–407.

265. Mann RS, Mulligan RC, Baltimore D. Construction of a retrovirus packaging mutant and its use to produce helper-free defective retrovirus. *Cell* 1983;32:871–879.

266. Maruyama M, Shibuya H, Harada H, et al. Evidence for aberrant activation of the interleukin-2 autocrine loop by HTLV-1-encoded p40x and T3/Tl complex triggering. *Cell* 1987;48:343–350.

267. Maurer B, Bannert H, Darai G, Flügel RM. Analysis of the primary structure of the long terminal repeat and the *gag* and *pol* genes of the human spumaretrovirus. *J Virol* 1988;62:1590–1597.

268. McClure MA, Johnson MS, Doolittle RF. Relocation of a protease-like gene segment between two retroviruses. *Proc Natl Acad Sci USA* 1987;84:2693–2697.

269. McClure MA, Johnson MS, Feng D, Doolittle, RF. Sequence comparisons of retroviral proteins: relative rates of change and general phylogeny. *Proc Natl Acad Sci USA* 1988;85: 2469–2473.

270. McCune JM, Rabin LB, Feinberg MB, et al. Endoproteolytic cleavage of gp160 is required for activation of human immunodeficiency virus. *Cell* 1988;53:55–67.

271. Meek TD, Dayton BD, Metcalf BW, et al. Human immunodeficiency virus 1 protease expressed in *Escherichia coli* behaves as a dimeric aspartic protease. *Proc Natl Acad Sci USA* 1989;86:1841–1845.

272. Meric C, Darlix JL, Spahr P. It is Rous sarcoma virus p12 and not p19 that binds tightly to Rous sarcoma virus RNA. *J Mol Biol* 1986;173:531–538.

273. Meric C, Goff SP. Characterization of Moloney murine leukemia virus mutants with single-amino-acid substitutions in the cys-his box of the nucleocapsid protein. *J Virol* 1989;63:1558–1568.

274. Meric C, Gouilloud E, Spahr P. Mutations in Rous sarcoma virus nucleocapsid protein p12 (NC): deletions of Cys-His boxes. *J Virol* 1988;62:3228–3233.

275. Meric C, Spahr P. Rous sarcoma virus nucleic acid binding protein p12 is necessary for viral 70S RNA dimer formation and packaging. *J Virol* 1986;60:450–459.

276. Mermer B, Malamy M, Coffin JM. Rous sarcoma virus contains sequences which permit expression of the *gag* gene in *Escherichia coli*. *Mol Cell Biol* 1983;3:1746–1758.

277. Michalides R, Wagenaar E, Weijers P. Rearrangements in the long terminal repeat of extra mouse mammary tumor proviruses in T-cell leukemias of mouse strain GR result in a novel enhancer-like structure. *Mol Cell Biol* 1985;5:823–830.

278. Miles BD, Robinson HL. High frequency transduction of c-erb B in avian leukosis virus-induced erythroblastosis. *J Virol* 1985;54:295–303.

279. Miller M, Jaskolski M, Mohana Rao JK, Leis J, Wlodawer A. Crystal structure of a retroviral protease proves relationship to aspartic protease family. *Nature* 1989;337:576–579.

280. Miller CK, Temin HM. Insertion of several different DNAs in reticuloendotheliosis virus strain T suppresses transformation by reducing the amount of sangenomic DNA. *J Virol* 1986;58:75–80.

281. Mizutkani T, Fuerst TR, Berger EA, Moss B. Binding region for human immunodeficiency virus (HIV) and epitopes for HIV-blocking monoclonal antibodies of the CD4 molecule defined by site-directed mutagenesis. *Proc Natl Acad Sci USA* 1988;85:9273–9277.

282. Mizutani S, Temin HM. Incorporation of noncomplementary deoxyribonucleotides at high frequencies by ribodeoxyvirus DNA polymerases and *Escherichia coli* DNA polymerase I. *Biochemistry* 1976;15:1516–1516.

283. Moore R, Dixon M, Smith R, Peters G, Dickson C. Complete nucleotide sequence of a milk-transmitted mouse mammary tumor virus: two frameshift suppression events are required for translation of *gag* and *pol*. *J Virol* 1987;61:480–490.

284. Mosca JD, Bednarik DP, Raj NBK, et al. Activation of human immunodeficiency virus by herpesvirus infection. 1. Identification of a region within the long terminal repeat that responds to a *trans*-acting factor encoded by herpes simplex virus 1. *Proc Natl Acad Sci USA* 1987;84:7408–7412.

285. Mosca JD, Bednarik DP, Raj NBK, et al. Herpes simplex virus type-1 can reactivate transcription of latent human immunodeficiency virus. *Nature* 1987;325:67–70.

286. Mount SM, Rubin GM. Complete nucleotide sequence of the Drosophila transposable element copia: homology between copia and retroviral proteins. *Mol Cell Biol* 1985;5:1630–1638.

287. Moss J, Helmer EP, LeBrice SFJ. Processing protease and reverse transcriptase from human immunodeficiency virus type 1 polyprotein in *Escherichia coli*. *J Virol* 1988;62:1433–1436.

288. Mowat M, Cheng A, Kimura N, Bernstein A, Benchimol S. Rearrangement of the cellular *p53* gene in erythroleukaemic cells transformed by Friend virus. *Nature* 1985;314:633–636.

289. Muesing MA, Smith DH, Cabradilla CD, Benton CV, Lasky LA, Capon DJ. Nucleic acid structure and expression of the human AIDS/lymphadenopathy retrovirus. *Nature* 1985;313: 450–458.

290. Muesing MA, Smith DH, Capon DJ. Regulation of mRNA accumulation by a human immunodeficiency virus *trans*-activator protein. *Cell* 1987;48:691–701.

291. Mullins JI, Brody DS, Binari RCJ, Cotter SM. Viral transduction of *c-myc* in naturally occurring feline leukaemias. *Nature* 1984;308:856–858.

292. Mullins JI, Chen CS, Hoover EA. Disease-specific and tissue-specific production of immunigrated feline leukaemia virus variant DNA in feline AIDS. *Nature* 1986;319:333–336.

293. Murphy JE, Goff SP. Construction and analysis of deletion mutations in the U5 region of Moloney murine leukemia virus: effects on RNA packaging and reverse transcription. *J Virol* 1988;63:319–327.

294. Myers G, Rabson AB, Josephs SF, Wong-Staal F. *Human retroviruses and AIDS 1988*. Los Alamos, New Mexico: Los Alamos National Laboratory, 1988.

295. Nabel G, Baltimore D. An inducible transcription factor activates human immunodeficiency virus expression in T-cells. *Nature* 1987;326:711–713.

296. Nabel GJ, Rice SA, Knipe DM, Baltimore D. Alternative mechanisms for activation of human immunodeficiency virus enhancer in T cells. *Science* 1988;239:1299–1302.

297. Navia MA, Fitzgerald PMD, McKeever BM, et al. Three-dimensional structure of aspartyl protease from human immunodeficiency virus HIV-1. *Nature* 1989;337:615–620.

298. Neil JC, Hughes D, McFarlane R, et al. Transduction and rear-

rangement of the *myc* gene by feline leukemia virus in naturally occurring T-cell leukaemias. *Nature* 308:814–820.

299. Nerenberg M, Hinrichs SH, Reynolds RK, Khoury G, Jay G. *The tat* gene of human T-lymphotropic virus type 1 induces mesenchymal tumors in transgenic mice. *Science* 1987;237: 1324–1329.

300. Nicolaisen-Sistrunk K, Kumar HPM, Fitting F, Grant CK, Elder JH. Natural feline leukemia virus variant escapes neutralization by a monoclonal antibody via an amino acid change outside the antibody-binding epitope. *J Virol* 1987;61:3410–3415.

301. Niedermann TMJ, Thielan BJ, Rabner L. Human immunodeficiency virus type 1 negative factor is a transcriptional silencer. *Proc Natl Acad Sci USA* 1989;86:1128–1132.

302. Nilsen TW, Maroney PA, Goodwin RG, et al. *c-erbB* activation in ALV-induced erythroblastosis: novel RNA processing and promoter insertion result in expression of an amino-truncated EGF receptor. *Cell* 1983;34:719–726.

303. Nishizawa M, Koyama T, Kawai S. Unusual features of the leader sequence of Rous sarcoma virus packaging mutant TK15. *J Virol* 1985;55:883–885.

304. Niwa O. Suppression of the hypomethylated Moloney leukemia virus genome in undifferentiated teratocarcinoma cells and inefficiency of transformation by a bacterial gene under control of the long terminal repeat. *Mol Cell Biol* 1985;5:2325–2331.

305. Niwa O, Yokota Y, Ishida N, Sugahara T. Independent mechanisms involved in suppression of the Moloney leukemia virus genome during differentiation of murine teratocarcinoma cells. *Cell* 1983;32:1105–1113.

306. Norton PA, Coffin JM. Bacterial β-galactosidase as a marker of Rous sarcoma virus gene expression and replication. *Mol Cell Biol* 1985;5:281–290.

307. Norton PA, Coffin JM. Characterization of Rous sarcoma virus sequences essential for viral gene expression. *J Virol* 1987;61:1171–1179.

308. Nusse R, Berns A. Cellular oncogene activation by insertion of retroviral DNA: genes identified by provirus tagging. In: Klein G, ed. *Cellular oncogene activation*. New York: Marcel Decker, 1988;95–119.

309. Nyborg JK, Dynan WS, Chen ISY, Wachsman W. Binding of host-cell factors to DNA sequences in the long terminal repeat of human T-cell leukemia virus type I: implications for viral gene expression. *Proc Natl Acad Sci USA* 1988;85:1457–1461.

310. Ohta M, Nyunoya H, Tanaka H, Okamoto T, Akagi T, Shimotohno K. Identification of a *cis*-regulatory element involved in accumulation of human T-cell leukemia virus type II genomic mRNA. *J Virol* 1988;62:4445–4451.

311. Okamoto T, Wong-Staal F. Demonstration of virus-specific transcriptional activator(s) in cells infected with HTLV-III by an *in vitro* cell-free system. *Cell* 1986;47:29–35.

312. Olsen JS, Swanstrom R. A new pathway in the generation of defective retrovirus DNA. *J Virol* 1985;56:779–789.

313. Omer CA, Faras AJ. Mechanism of release of the avian retrovirus RNA·Trp primer molecule from viral DNA by ribonuclease H during reverse transcription. *Cell* 1982;30:797–805.

314. Omer CA, Pogue-Geile K, Gunthaka B, Staskis KA, Faras AJ. Involvement of directly repeated sequences in the generation of deletions of the avian sarcoma virus *src* gene. *J Virol* 1983;47:380–382.

315. Ono M. Molecular cloning and long terminal repeat sequences of human endogenous retrovirus genes related to types A and B retrovirus genes. *J Virol* 1986;58:937–944.

316. Ostrove JM, Leonard J, Weck KE, Rabson AB, Gendelman HE. Activation of the human immunodeficiency virus by herpes simplex virus type 1. *J Virol* 1987;61:3726–3732.

317. Ott C, Boone LR, Koh CK, Tennant RW, Yang WK. Nucleotide sequences of gag-pol regions that determine the *Fv-1* host range property of BALB/c N-tropic and B-tropic murine leukemia viruses. *J Virol* 1983;48:776–784.

318. Overbaugh J, Donahue PR, Quackenbush SL, Hoover EA, Mullins JI. Molecular cloning of a feline leukemia virus that induces fatal immunodeficiency disease in cats. *Science* 1988;239:906–910.

319. Panet A, Bahimore D. Characterization of endonuclease activities on Moloney murine leukemia virus and its replication defective mutants. *J Virol* 1977;61:1756–1760.

320. Panganiban AT, Flore D. Ordered interstrand and intrastrand DNA transfer during reverse transcription. *Science* 1988;241:1064–1069.

321. Panganiban AT, Temin HM. The terminal nucleotides of retrovirus DNA are required for integration but not virus production. *Nature* 1983;306:155–160.

322. Panganiban AT, Temin HM. Circles with two tandem LTR's are precursors to integrated retrovirus DNA. *Cell* 1984;36:673–679.

323. Panganiban AT, Temin HM. The retrovirus *pol* gene encodes a product required for DNA integration: Identification of a retrovirus *int* locus. *Proc Natl Acad Sci USA* 1984;81:7885–7889.

324. Papsidero LD, Sheu M, Ruscetti FW. Human immunodeficiency virus type 1-neutralizing monoclonal antibodies which react with p17 core protein. characterization and epitope mapping. *J Virol* 1988;63:267–272.

325. Parvin JD, Moscona A, Pan WT, Leider JM, Palese P. Measurement of the mutation rates of animal viruses: Influenza A and poliovirus type 1. *J Virol* 1986;59:377–383.

326. Paskalis H, Felber BK, Pavlakis GN. *Cis*-acting sequences responsible for the transcriptional activation of human T-cell leukemia virus type I constitute a conditional enhancer. *Proc Natl Acad Sci USA* 1986;83:6558–6562.

327. Pearl LH, Taylor WR. Sequence specificity of retroviral proteases. *Nature* 1987;328:482–483.

328. Pepinsky RB, Vogt VM. Fine-structure analyses of lipid–protein and protein–protein interactions of gag protein p19 of the avian sarcoma and leukemia viruses by cyanogen bromide mapping. *J Virol* 1984;52:145–153.

329. Perez LG, Davis GL, Hunter E. Mutants of the Rous sarcoma virus envelope glycoprotein that lack the transmembrane anchor and cytoplasmic domains: analysis of intracellular transport and assembly into virions. *J Virol* 1987;61:2981–2988.

330. Perez LG, Hunter E. Mutations within procolytic cleavage site of the Rous sarcoma virus glycoprotein that block processing to gp85 and gp37. *J Virol* 1987;61:1609–1614.

331. Peters GG, Hu J. Reverse transcriptase as the major determinant for selective packaging of tRNA's into avian sarcoma virus particles. *J Virol* 1980;36:692–700.

332. Peters G, Placzek M, Brookes S, Kozak C, Smith R, Dickson C. Characterization, chromosomic assignment, segregation analysis of endogenous proviral units of mouse mammary tumor virus. *J Virol* 1986;59:535–544.

333. Petersen RB, Hackett PB. Characterization of ribosome binding on Rous sarcoma virus RNA *in vitro*. *J Virol* 1985;56:683–690.

334. Pillemer EA, Kooistra DA, Witte ON, Weissman IL. Monoclonal antibody to the amino-terminal L sequence of murine leukemia virus glycosylated gag poly proteins demonstrates their unusual orientation in the cell membrane. *J Virol* 1986;57:413–421.

335. Pinter A, Honnen WJ. O-linked glycosylation of retroviral envelope gene products. *J Virol* 1988;62:1016–1021

336. Portis JL, Atec FJ, Evans LH. Infectious entry of murine retroviruses into mouse cells: evidence of a postadsorption step inhibited by acidic pH. *J Virol* 1983;55:806–812.

337. Power MD, Marx PA, Bryant ML, Gardener MB, Barr PJ, Luciw PA. Nucleotide sequence of SRV-1, a type D simian acquired immune deficiency syndrome retrovirus. *Science* 1986;231:1567–1572.

338. Prats AC, Sarih L, Gabus C, Litvak S, Keith G, Darlix J. Small finger protein of avian and murine retroviruses has nucleic acid annealing activity and positions the replication primer tRNA onto genomic RNA. *EMBO J* 1988;7:1135–1139

339. Preston BD, Poiesz BJ, Loeb LA. Fidelity of HIV-1 reverse transcriptase. *Science* 1988;242:1168–1171

340. Ravretskis J. Expression of the protein product of the mouse mammary tumor virus long terminal repeat gene in phorbol ester-treated mouse T-cell leukemia cells. *J Virol* 1986;58:441–449.

341. Raines MA, Maihle NJ, Moscovici C, Gutfreden L, Kung H. Molecular characterization of three erbB transducing viruses

generated during avian leukosis virus-induced erythroleukemia: extensive internal deletion near the kinase domain activates the fibrosarcoma- and hemangioma-induced potential of erbB. *J Virol* 1988;62:2437–2443.

342. Raitray AJ, Champoux JJ. The role of Moloney murine leukemia virus RNase H activity in the formation of plus-strand primers. *J Virol* 1987;61:2843–2851.

343. Rein A. Interference grouping of murine leukemia viruses: a distinct receptor for MCF-recombinant viruses in mouse cells. *Virology* 1982;120:251–257.

344. Rein A, McClure MR, Rice NR, Luftig RB, Schultz AM. Myristylation site in Pr65gag is essential for virus particle formation by Moloney murine leukemia virus. *Proc Natl Acad Sci USA* 1986;83:7246–7250.

345. Repaske R, Steele PE, O'Neill RR, Rabson AB, Martin MA. Nucleotide sequence of a full-length human endogenous retroviral segment. *J Virol* 1985;54:764–772.

346. Rhee SS, Hunter E. Myristylation is required for intracellular transport but not for assembly of D-type retrovirus capsids. *J Virol* 1987;61:1045–1053.

347. Richardson NE, Brown NR, Hussey RE, et al. Binding site for human immunodeficiency virus coat protein gp120 is located in the NH$_2$-terminal region of T4 (CD4) and requires the intact variable-region-like domain. *Proc Natl Acad Sci USA* 1988;85:6102–6106.

348. Rimsky L, Hauber J, Dukovich M, et al. Functional replacement of the HIV-1 rev protein by the HTLV-1 rex protein. *Nature* 1988;335:738–740.

349. Roberts JD, Bebenek K, Kunkel TA. The accuracy of reverse transcriptase from HIV-1. *Science* 1988;242:1171–1173.

350. Roberts JD, Preston BD, Johnston LA, Soni A, Loeb LA, Kunkel T. Fidelity of two retroviral reverse transcriptases during DNA-dependent DNA synthesis *in vitro*. *Mol Cell Biol* 1989;9:469–476.

351. Robinson HL, Bhatt BM, Talcliffe PN, Coffin JM. At least two regions of the viral genome determine the oncogenic potential of avian leukosis viruses. *Proc Natl Acad Sci USA* 1982;79:1225–1229.

352. Robinson HL, Gagnon GC. Patterns of proviral insertion in avian leukosis virus-induced lymphomas. *J Virol* 1986;57:28–36.

353. Robinson HL, Jensen L, Coffin JM. Sequences outside of the long terminal repeat determine the lymphomagenic potential of Rous-associated virus type 1. *J Virol* 1985;55:752–759.

354. Robinson HL, Miles BD. Avian leukosis virus induced osteopetrosis is associated with the persistent synthesis of viral DNA. *Virology* 1985;141:130–143.

355. Robinson HL, Reinsch SS, Shank PR. Sequences near the 5′ long terminal repeat of avian leukosis viruses determine the ability to induce osteopetrosis. *J Virol* 1986;59:45–49.

356. Rohdewohld H, Weiher H, Reik W, Jaenisch R, Breindl M. Retrovirus integration and chromatin structure: Moloney murine leukemia proviral integration sites map near DNase I-hypersensitive sites. *J Virol* 1987;61:336–343.

357. Rommelaere J, Domis-Keller H, Hopkins N. RNA sequencing provides evidence for allelism of determinants of the N-, B-, or NB-tropism of murine leukemia viruses. *Cell* 1979;16:43–50.

358. Rosen CA, Haseltine WA, Lenz J, Ruprecht R, Cloyd MW. Tissue selectivity of murine leukemia virus infection is determined by long terminal repeat sequences. *J Virol* 1985;55:862–866.

359. Rosen CA, Park B, Sodroski JG, Haseltine WA. Multiple sequence elements are required for regulation of human T-cell leukemia virus gene expression. *Proc Natl Acad Sci USA* 1987;84:4919–4923.

360. Rosen CA, Sodroski JG, Goh WC, Dayton AI, Lippke J, Haseltine WA. Post-transcriptional regulation accounts for the *trans*-activation of the human T-lymphotropic virus type III. *Nature* 1986;319:555–559.

361. Rosen CA, Sodroski JG, Haseltine WA. Location of *cis*-acting regulatory sequences in human T cell lymphotropic virus type III (HTLV III/LAV) long terminal repeat. *Cell* 1985;41:813–823.

362. Rosen CA, Terwilliger E, Dayton A, Sodroski JG, Haseltine WA. Intragenic *cis*-acting art gene-responsive sequences of the human immunodeficiency virus. *Proc Natl Acad Sci USA* 1988;85:2071–2075.

363. Rosenblatt JD, Cann AJ, Slamon DJ, et al. HTLV-II transactivation is regulated by the overlapping *tax/rex* nonstructural genes. *Science* 1988;240:916–918.

364. Roth MJ, Tanese N, Goff SP. Gene product of Moloney murine leukemia virus required for proviral integration is a DNA-binding protein. *J Mol Biol* 1989;209:131–139.

365. Ruben S, Perkat H, Tan T, et al. Cellular transcription factors and regulation of IL-2 receptor gene expression by HTLV-I *tax* gene product. *Science* 1988;241:89–92.

366. Ryden TA, Bestman K. Avian retroviral long terminal repeats bind CCAAT/enhancer-binding protein. *Mol Cell Biol* 1989;9:1155–1164.

367. Schiff LA, Nibert ML, Fields BN. Characterization of a zinc blotting technique: evidence that a retroviral gag protein binds zinc. *Proc Natl Acad Sci USA* 1988;85:4195–4199.

368. Schultz AM, Rein A. Unmyristylated Moloney murine leukemia virus Pr65gag is excluded from virus assembly and maturation events. *J Virol* 1989;63:2370–2372.

369. Sealey L, Chalkley R. At least two nuclear proteins bind specifically to the Rous sarcoma virus long terminal repeat enhancer. *Mol Cell Biol* 1987;7:787–798.

370. Seiki M, Hattori S, Hirayama Y, Yoshida M. Human adult T-cell leukemia virus: complete nucleotide sequence of the provirus genome integrated in leukemia cell DNA. *Proc Natl Acad Sci USA* 1983;80:3618–3622.

371. Seiki M, Hikikoshi A, Taniguchi T, Yoshida M. Expression of the *pX* gene of HTLV-I: general splicing mechanism in the HTLV family. *Science* 1985;228:1532–1534.

372. Seiki M, Inoue J, Kidaka M, Yoshida M. *Two cis*-acting elements responsible for post-transcriptional *trans*-regulation of gene expression of human T-cell leukemia virus type I. *Proc Natl Acad Sci USA* 1988;85:7124–7128.

373. Selten G, Cuypers HT, Zijlstra M, Melief C, Berns A. Involvement of *c-myc* in Mul.V-induced T-cell lymphomas in mice. Frequency and mechanism of action. *EMBO J* 1984;3:3215–3222.

374. Seperack PK., Strobel MC, Corrow DJ, Jenkins NA, Copeland NG. Somatic and germ-line reverse mutation rates of the retrovirus-induced dilute coat-color mutation of DBA mice. *Proc Natl Acad Sci USA* 1988;85:189–192.

375. Seto E, Yen TSB, Peterlin BM, Ou J. *Trans*-activation of the human immunodeficiency virus long terminal repeat by the hepatitis B virus X protein. *Proc Natl Acad Sci USA* 1988;85:8286–8290.

376. Shih C, Stoye JP, Coffin JM. Highly preferred targets for retrovirus integration. *Cell* 1988;53:531–537.

377. Shimotohno K, Takami M, Teruuchi T, Miwa M. Requirement of multiple copies of a 21-nucleotide sequence in the U3 regions of human T-cell leukemia virus type I and type II long terminal repeats for *trans*-acting activation of transcription. *Proc Natl Acad Sci USA* 1986;83:8112–8116.

378. Shimotohno K, Temin HM. Spontaneous variation and synthesis in the U3 region of the long terminal repeat of an avian retrovirus. *J Virol* 1982;41:163–171.

379. Shoemaker CS, Goff S, Gilboa E, Paskind M, Mitra SW, Baltimore D. Structure of a cloned circular Moloney murine leukemia virus molecule containing an inverted segment: implications for retrovirus integration. *Proc Natl Acad Sci USA* 1980;77:3932–3936.

380. Short MK, Okenquist SA, Lenz J. Correlation of leukemogenic potential of murine retroviruses with transcriptional tissue preference of the viral long terminal repeats. *J Virol* 1987;61:1067–1072.

381. Siddiqui A, Gaynor R, Srinivasan A, Mapoles J, Farr RW. *trans*-activation of viral enhancers including long terminal repeat of the human immunodeficiency virus by the hepatitis B virus X protein. *Virology* 1989;169:479–484.

382. Siekevitz M, Feinberg MB, Holbrook N, Wong-Staal F, Greene WC. Activation of interleukin 2 and interleukin 2 receptor (Tac) promoter expression by the *trans*-activator (*tat*) gene product

of human T-cell leukemia virus, type I. *Proc Natl Acad Sci USA* 1987;84:5386–5393.
383. Skalka AM. Integrative recombination in retroviruses. In: Kuchelapati R, Smith GR, eds. *Genetic recombination*. Washington, DC: ASM, 1988;701–724.
384. Skalka AM. Retroviral proteases: first glimpses at the anatomy of a processing machine. *Cell* 1989;56:911–913.
385. Skalka AM, Leis J. Retroviral DNA integration. *Bioassays* 1984;1:206–210.
386. Smith DB, Inglis SC. The mutation rate and variability of eukaryotic viruses: an analytical review. *J Gen Virol* 1987;68:2729–2740.
387. Smith TF, Srinivasan A, Schochetman G, Marcus M, Myers G. The phylogenetic history of immunodeficiency viruses. *Nature* 1988;333:573–575.
388. Sodroski J, Goh WC, Rosen C, Campbell K, Haseltine WA. Role of the HTLV-III/LAV envelope in syncytium formation and cytopathicity. *Nature* 1986;322:470–474.
389. Sodroski J, Goh WC, Rosen C, Dayton A, Terwilleger E, Haseltine W. A second post-transcriptional *trans*-activator gene required for HTLV-III replication. *Nature* 1986;321:412–416.
390. Sodroski J, Rosen C, Goh WC, Haseltine W. A transcriptional activator protein encoded by the x-lor region of the human T-cell leukemia virus. *Science* 1985;228:1430–1434.
391. Somasundaram M, Robinson HL. A major mechanism of human immunodeficiency virus-induced cell killing does not involve cell fusion. *J Virol* 1987;61:3114–3119.
392. Somasundaram M, Robinson HL. Unexpectedly high levels of HIV-1 RNA and protein synthesis in a cytocidal infection. *Science* 1988;242:1554–1557.
393. Songo P, Alizon M, Staskus K, et al. Nucleotide sequence of the visna lentivirus: relationship to the AIDS virus. *Cell* 1985;42:369–382.
394. Sonigo P, Barker C, Hunter E, Wain-Hobson S. Nucleotide sequence of Mason–Pfizer monkey virus: an immunosuppressive D-type retrovirus. *Cell* 1986;45:375–385.
395. Sorge J, Ricel W, Hughes SH. *cis*-Acting packaging locus in the 115-nucleotide direct repeat of Rous sarcoma virus. *J Virol* 1983;48:667–675.
396. Speck NA, Baltimore D. Six distinct nuclear factors interact with the 75-base-pair repeat of the Moloney murine leukemia virus enhancer. *Mol Cell Biol* 1987;7:1101–1110.
397. Spence SE, Gilbert LJ, Swing DA, Copeland NG, Jenkins NA. Spontaneous germ line virus infection and retroviral insertional mutagenesis in eighteen transgenic Srev lines of mice. *Mol Cell Biol* 1989;9:177–184.
398. Spiro C, Li J, Beswick RK, Kabat D. An enhancer sequence instability that diversifies the cell repertoire for expression of a murine leukemia virus. *Virology* 1988;164:350–361.
399. Stavenhagen JB, Robins DM. An ancient provirus has imposed androgen regulation on the adjacent mouse sex-limited protein gene. *Cell* 1988;55:247–254.
400. Steele PE, Martin MA, Rabson AB, Bryan T, O'Brien SJ. Amplification and chromosomal dispersion of human endogenous retroviral sequences. *J Virol* 1986;59:545–550.
401. Stein BS, Gowda SD, Lifson JD, Penhallow RC, Bensch KG, Engelman EG. pH-independent HIV entry into CD4-positive T cells via virus envelope fusion to the plasma membrane. *Cell* 1987;49:659–669.
402. Steinhauer DA, de la Torre JC, Meier E, Holland JJ. Extreme heterogeneity in populations of vesicular stomatitis virus. *J Virol* 1989;63:2072–2080.
403. Steinhauer DA, Holland JJ. Direct method for quantitation of extreme polymerase error frequencies at selected single base sites in viral RNA. *J Virol* 1987;57:219–228.
404. Stewart CL, Stuhlmann H, Jähner D, Jaenisch R. *De novo* methylation, expression, and infectivity of retroviral genomes introduced into embryonal carcinoma cells. *Proc Natl Acad Sci USA* 1982;79:4098–4102.
405. Stocking C, Kollek R, Bergholz U, Ostertag W. Long terminal repeat sequences impart hematopoietic transformation properties to the myeloproliferative sarcoma virus. *Proc Natl Acad Sci USA* 1985;82:5746–5750.
406. Stocking C Kollek R, Bergholz U, Ostertag W. Point mutations in the U3 region of the long terminal repeat of Moloney murine leukemia virus determine disease specificity of the myeloproliferative sarcoma virus. *Virology* 1986;153:145–149.
407. Stoltzfus CM, Chiang L, Cripe TP, Turek LP. Efficient transformation by Prague A Rous sarcoma virus plasmid DNA requires the presence of *cis*-acting regions within the *gag* gene. *J Virol* 1987;61:3401–3409.
408. Stoltzfus CM, Dane RW. Accumulation of spliced avian retroviral RNA is inhibited in 3′-adenosyl methionine depleted chicken embryo fibroblasts. *J Virol* 1982;42:918–931.
409. Stoltzfus CM, Fogarty CJ. Multiple regions in the Rous sarcoma virus *src* gene intron act in cis to affect the accumulation of unspliced RNA. *J Virol* 1989;63:1669–1676.
410. Stoltzfus CM, Lowmiller SK, Borberich SL. Noncoding region between the *env* and *src* genes of Rous sarcoma virus influences splicing efficiency of the *src* gene 3′ splice site. *J Virol* 1987;61:177–184.
411. Stoye JP, Coffin JM. Endogenous viruses. In: Weiss R, Teich N, Varmus H, Coffin J, eds. *RNA tumor viruses*. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory, 1985;357–404.
412. Stoye JP, Coffin JM. The four classes of endogenous murine leukemia virus: structural relationships and potential for recombination. *J Virol* 1987;61:2659–2669.
413. Stoye JP, Fenner S, Greenoak GE, Moran C, Coffin JM. Role of endogenous retroviruses as mutagens: the hairless mutation of mice. *Cell* 1988;54:383–391.
414. Strauss EG, Rice CM, Strauss JH. Sequence coding for the alphavirus nonstructural proteins is interrupted by an opal codon. *Proc Natl Acad Sci USA* 1983;80:5271–5275.
415. Stuhlmann H, Jaenisch R, Mulligan RC. Construction and properties of replication-competent murine retroviral vectors encoding methotrexate resistance. *Mol Cell Biol* 1989;9:100–108.
416. Svoboda J, Doerak M, Guntaka R, Geryk J. Transmission of (LTR-v-src-LTR) without recombination with a helper virus. *Virology* 1986;153:314–318.
417. Swain A, Coffin JM. Polyadenylation at correct sites is not required for retrovirus replication or genome encapsidation. *J Virol* 1989;*(In press)*.
418. Swanstrom R, Parker RC, Varmus HE, Bishop JM. Transduction of a cellular oncogene: the genesis of Rous sarcoma virus. *Proc Natl Acad Sci USA* 1983;80:2519–2523.
419. Szurek PF, Yuen PH, Jerzy R, Wong PKY. Identification of point mutations in the envelope gene of Moloney murine leukemia virus TB temperature-sensitive paralytogenic mutants *ts1*: molecular determinants for neurovirulence. *J Virol* 1988;62:357–360.
420. Taketo M, Gilboa E, Sherman ML. Isolation of embryonal carcinoma cell lines that express integrated recombinant genes flanked by the Moloney murine leukemia virus long terminal repeat. *Proc Natl Acad Sci USA* 1985;82:2422–2426.
421. Taketo M, Tanaka M. A cellular enhancer of retrovirus gene expression in embryonal carcinoma cells. *Proc Natl Acad Sci USA* 1987;84:3748–3752.
422. Tanese N, Goff SP. Domain structure of the Moloney murine leukemia virus reverse transcriptase: mutational analysis and separate expression of the DNA polymerase and RNase H activities. *Proc Natl Acad Sci USA* 1988;85:1777–1781.
423. Teich N. Taxonomy of retroviruses. In: Weiss R, Teich N, Varmus H, Coffin J, eds. *RNA tumor viruses*. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory, 1982;25–208.
424. Teich N. Taxonomy of retroviruses. In: Weiss R, Teich N, Varmus H, Coffin J, eds. *RNA tumor viruses, 2nd edition*. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory, 1985;1–16.
425. Teich N, Weiss RA, Martin GR, Lowy DR. Virus infection of murine teratocarcinoma stem cell lines. *Cell* 1977;12:973–982.
426. Teich N, Wyke J, Kaplan P. Pathogenesis of retrovirus-induced disease. In: Weiss R, Teich N, Varmus H, Coffin J, eds. *RNA tumor viruses, 2nd edition*. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory, 1985;787–248.
427. Teich N, Wyke J, Mak T, Bernstein A, Hardy W. Pathogenesis of retrovirus-induced disease. In: Weiss R, Teich N, Varmus

H. Coffin J, eds. *RNA tumor Viruses*. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory, 1982:785–998.
428. Temin HM. Nature of the provirus of Rous sarcoma. *Natl Cancer Inst Monogr* 1964;17:557–570.
429. Temin HM. Formation and activation of the provirus of RNA sarcoma virus. In: Barry RD, Mahy BW, eds. *The biology of large RNA viruses*. London: Academic Press, 1979:233–244.
430. Temin HM. Reverse transcription in the eukaryotic genome: retroviruses, pararetroviruses, retrotransposons, and retrotranscripts. *Mol Biol Evol* 1985;6:455–468.
431. Terry R, Sobis DA, Katzman M, Cobrnik D, Lois J, Skalka AM. Properties of avian sarcoma–leukosis virus pp32-related pol-endonucleases produced in *Escherichia coli*. *J Virol* 1988;62:2358–2365.
432. Tevethia MJ, Spector DJ. Heterologous transactivation among viruses. *Prog Med Virol* 1989; (*in press*).
433. Toohey MG, Jones KA. *In vitro* formation of short RNA polymerase II transcripts can terminate within the HIV-1 and HIV-2 promoter-proximal downstream regions. *Genes Dev* 1989;3:265–282.
434. Tsichlis PN, Coffin JM. Recombinants between endogenous and exogenous avian tumor viruses: role of the C region and other portions of the genome in the control of replication and transformation. *J Virol* 1980;33:238–249.
435. Tsichlis PN, Strauss PG, Lohse MA. Concerted DNA rearrangements in Moloney murine leukemia virus-induced thymomas: a potential synergistic relationship in oncogenesis. *J Virol* 1985;56:258–267.
436. Twu J, Robinson WS. Hepatitis B virus *X* gene can transactivate heterologous viral sequences. *Proc Natl Acad Sci USA* 1989;86:2046–2050.
437. Van Beveren C, Coffin JM, Hughes S. Appendixes. In: Weiss R, Teich N, Varmus H, Coffin J, eds. *RNA tumor viruses, 2nd edition*. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory, 1985:559–232.
438. van Klaveren P, Bentvelzen P. Transactivating potential of the 3′ open reading frame of Murine mammary tumor virus. *J Virol* 1988;62:4410–4413.
439. van Lohuizen M, Verbeek S, Krimpenfort P, et al. Predisposition to lymphomagenesis in pim-1 transgenic mice: cooperation with *c-myc* and *N-myc* in murine leukemia virus-induced tumors. *Cell* 1989;56:673–682.
440. Varmus HE. Reverse transcription. *Sci Am* 1987;257:56–66.
441. Varmus HE. Regulation of HIV and HTLV gene expression. *Genes Dev* 1988;2:1055–1062.
442. Varmus H. Retroviruses. *Science* 1988;240:1427–1435.
443. Varmus HE. Reverse transcription in bacteria. *Cell* 1989;56:721–724.
444. Varmus H, Brown P. Retroviruses. In: Howe M, Berg D, eds. *Mobile DNA*. Washington, DC: ASM, 1989;53–108.
445. Varmus HE, Quintrell NE, Ortiz S. Retroviruses as mutagens: insertion and excision of a non-transforming provirus alters expression of a resident transforming provirus. *Cell* 1981; 25:23–26.
446. Varmus HE, Swanstrom R. Replication of Retroviruses. In: Weiss R, Teich N, Varmus H, Coffin J, eds. *RNA tumor viruses*. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory, 1982:369–512.
447. Varmus HE, Swanstrom R. Replication of retroviruses. In: Weiss R, Teich N, Varmus H, Coffin J, eds. *RNA tumor viruses, 2nd edition*. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory, 1985:74–134.
448. Vijaya S, Steffen DL, Kozak C, Robinson HL. Dsi-1, a region with frequent proviral insertions in Moloney murine leukemia virus-induced rat thymomas. *J Virol* 1987;61:1164–1170.
449. Vijaya S, Steffen DL, Robinson HL. Acceptor sites for retroviral integrations map near DNase I-hypersensitive sites in chromatin. *J Virol* 1986;60:683–692.
450. Villemur R, Momczak Y, Rassart E, Kozak C, Jolicoeur P. Identification of a new common provirus integration site in Gross passage A murine leukemia virus-induced mouse thymoma DNA. *Mol Cell Biol* 1987;7:512–522.
451. Vogel SHH, Reynolds RK, Luciw PA, Jay G. The HIV *tat* gene induces dermal lesions resembling Kaposi's sarcoma in transgenic mice. *Nature* 1988;335:606–611.
452. Voynow SL, Coffin JM. Evolutionary variants of Rous Sarcoma virus: large deletion mutants do not result from homologous recombination. *J Virol* 1985;55:67–78.
453. Voynow SL, Coffin JM. Truncated gag-related proteins are produced by large deletion mutants of Rous sarcoma virus and form virus particles. *J Virol* 1985;55:79–85.
454. Voytas DF, Ausubel FM. A copia-like transposable element family in *Arabidopsis thaliana*. *Nature* 1988;336:242–244.
455. Wachsman W, Golde DW, Temple PA, Orr EC, Clark SC, Chen ISY. HTLV *x*-gene product: requirement for the env nucleotide initiation codon. *Science* 1985;228:1534–1537.
456. Watanabe S, Temin HM. Encapsidation sequences for spleen necrosis virus, an avian retrovirus, are between the 5′ long terminal repeat and the start of the *gag* gene. *Proc Natl Acad Sci USA* 1982;79:5986–5990.
457. Weber IT, Miller M, Jaskolski M, Leis J, Skalka AM, Wlodawer A. Molecular modeling of the HIV-1 protease and its substrate binding site. *Science* 1989;243:928–931.
458. Weiher H, Barklis E, Ostertag W, Jaenisch R. Two distinct sequence elements mediate retroviral gene expression in embryonal carcinoma cells. *J Virol* 1987;61:218–225.
459. Weiss RA. Experimental biology and assay of RNA tumor viruses. In: *RNA tumor viruses* Weiss R, Teich N, Varmus H, Coffin J, eds. Cold Spring Harbor New York: Cold Spring Harbor Laboratory, 1982:205–206.
460. Weiss RA. Human T-cell retroviruses. In: Weiss R, Teich N, Varmus H, Coffin J, eds. *RNA tumor viruses, 2nd edition*. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory, 1985:405–486.
461. Weiss R, Teich N, Varmus H, Coffin J, eds. *RNA tumor viruses*. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory, 1982.
462. Weiss R, Teich N, Varmus H, Coffin J, eds. *RNA tumor viruses, 2nd edition*. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory, 1985.
463. Weiler SK, Joy AE, Temin HM. Correlation between cell killing and massive second round superinfection by members of some subgroups of avian leukosis virus. *J Virol* 1980;33:494–506.
464. Willey RL, Smith DH, Lasky LA, et al. *In vitro* mutagenesis identifies a region within the envelope gene of the human immunodeficiency virus that is critical for infectivity. *J Virol* 1988;62:139–147.
465. Williams DA, Orkin SH, Mulligan RC. Retrovirus-mediated transfer of human adenosine deaminase gene sequences into cells in culture and into murine hematopoietic cells *in vivo*. *Proc Natl Acad Sci USA* 1986;83:2566–2571.
466. Wills JW, Shriniva RV, Hunter E. Mutations of the Rous sarcoma virus *env* gene that affect the transport and subcellular location of the glycoprotein products. *J Cell Biol* 1984;99:2011–2023.
467. Wilson W, Braddock M, Adams SE, Rathjen PD, Kingsman SM, Kingsman AJ. HIV expression strategies: ribosomal frameshifting is directed by a short sequence in both mammalian and yeast systems. *Cell* 1988;55:1159–1169.
468. Wilson W, Mahm MH, Mellor J, Kingsman AJ, Kingsman SM. Expression strategies of the yeast retrotransposon Ty: a short sequence directs ribosomal frameshifting. *Nucleic Acids Res* 1986;14:7001–7016.
469. Wilson C, Reitz MS, Okayama H, Eiden MV. Formation of infectious hybrid virions with Gibbon ape leukemia virus and human T-cell leukemia virus retroviral envelope glycoproteins and the gag and pol proteins of Moloney murine leukemia virus. *J Virol* 1989;63:2374–2378.
470. Wolff L, Rascetti S. The spleen focus-forming virus (SFFV) envelope gene, when introduced into mice in the absence of other SFFV genes, induces acute erythroleukemia. *J Virol* 1988;62:2158–2163.
471. Wong PKY, Knapp C, Yuen CKPH, Soong MM, Sachary JF, Tompkins WAF. *trf*, a paralytogenic mutant of Moloney murine leukemia virus TB, has an enhanced ability to replicate in

the central nervous system and primary nerve cell culture. *J Virol* 1985;55:760–767.
472. Wong-Staal F, Gallo RC. Human T-lymphotropic retroviruses *Nature* 1985;317:395–403
473. Yokoyama S, Gojobori T. Molecular evolution and phylogeny of the human AIDS viruses LAV, HTLV-III, and ARV. *J Mol Evol* 1987;23:330–336.
474. Yoshinaka Y, Katoh I, Copeland TD, Oroszlan S. Murine leukemia protease is encoded by the *gag-pol* gene and is synthesized through suppression of an amber termination codon. *Proc Natl Acad Sci USA* 1985;82:1618–1622.
475. Yoshinaka Y, Katoh I, Copeland TD, Oroszlan S. Translational readthrough of an amber termination codon during synthesis of feline leukemia virus protease. *J Virol* 1985;55:870–873.
476. Cann AJ, Chen ISY. Human T-Cell Leukemia Virus Types I and II. In: Fields BN, Knipe DM, et al., eds. *Virology*, 2nd Ed. New York: Raven Press, 1990:1501–1527.
477. Hirsch MS, Curran J. Human Immunodeficiency Viruses. In: Fields BN, Knipe DM, et al., eds. *Virology*, 2nd Ed. New York: Raven Press, 1990:1545–1570.
478. Narayan O, Clements JE. Lentiviruses. In: Fields BN, Knipe DM, et al., eds. *Virology*, 2nd Ed. New York: Raven Press, 1990:1571–1589.

1: Protein Seq Data Anal. 1990 Jul;3(3):267-71.

Links

Calculating percent identity between protein or DNA sequences with a word processor.

Ladror US.

Department of Biological Chemistry and Structure, University of Health Sciences, Chicago Medical School, IL 60064.

Two macros, to calculate percentage identity between protein or DNA sequences using the Microsoft Word word processor, are described. The user prepares an alignment file of multiple sequences which is used by the macros to calculate number of matches, number of mismatches, total number of compared positions, and the percent identity. The macros are especially useful when alignment of multiple sequences is possible only by eye.

PMID: 2402483 [PubMed - indexed for MEDLINE]

1: Science. 1986 May 16;232(4752):854-8.

Cloning of a cDNA for a T cell-specific serine protease from a cytotoxic T lymphocyte.

Gershenfeld HK, Weissman IL.

A new serine protease was encoded by a clone isolated from a murine cytotoxic T-lymphocyte complementary DNA library by an RNA-hybridization competition protocol. Complementary transcripts were detected in cytotoxic T lymphocytes, spleen cells from nude mice, a rat natural killer cell leukemia, and in two of eight T-helper clones (both cytotoxic), but not in normal mouse kidney, liver, spleen, or thymus, nor in several tested T- and B-cell tumors. T-cell activation with concanavalin A plus interleukin-2 induced spleen cells to express this gene with kinetics correlating with the acquisition of cytolytic capacity. The nucleotide sequence of this gene encoded an amino acid sequence of approximately 25,700 daltons, with 25 to 35 percent identity to members of the serine protease family. The active site "charge-relay" residues (His57, Asp102, and Ser195 of the chymotrypsin numbering system) are conserved, as well as the trypsin-specific Asp (position 189 in trypsin). A Southern blot analysis indicated that this gene is conserved in humans, mouse, and chicken. This serine protease may have a role in lymphocyte lysis and a "lytic cascade."

PMID: 2422755 [PubMed - indexed for MEDLINE]